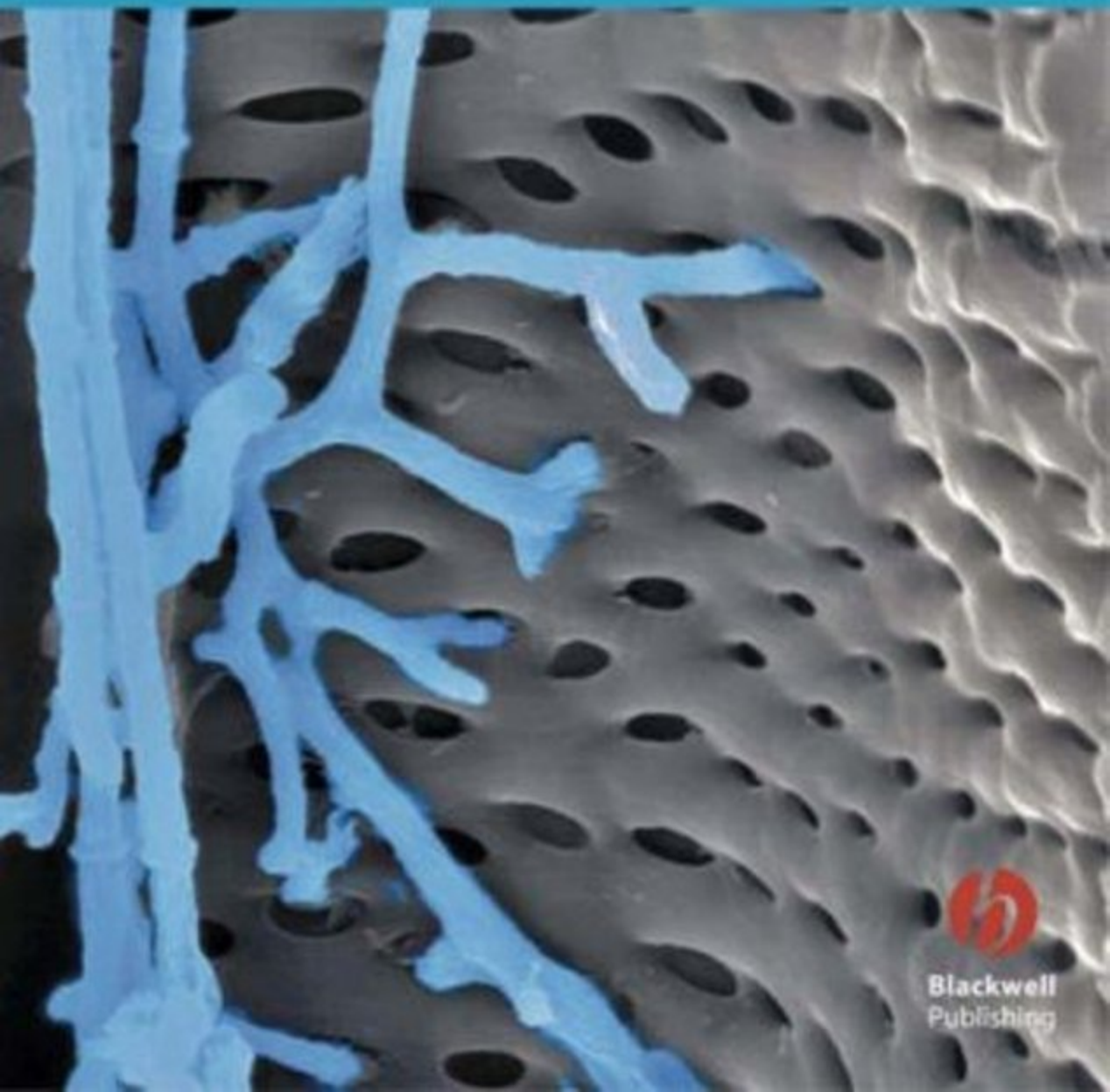


Biomass Recalcitrance

Deconstructing the Plant Cell Wall for Bioenergy

Edited by Michael E. Himmel



Blackwell
Publishing

Biomass Recalcitrance

Deconstructing the Plant Cell Wall for Bioenergy

Biomass Recalcitrance

Deconstructing the Plant Cell Wall for Bioenergy

Edited by

Dr Michael E. Himmel

Principal Scientist, National Renewable
Energy Laboratory, Golden, CO, USA

This edition first published 2008
© 2008 Blackwell Publishing Ltd

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing programme has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial office

9600 Garsington Road, Oxford, OX4 2DQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of the author to be identified as the author of this Work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Biomass recalcitrance : deconstructing the plant cell wall for bioenergy/edited by Dr. Michael E. Himmel.
p. cm.

Includes bibliographical references and index.

ISBN-13: 978-1-4051-6360-6 (hardback : acid-free paper)

ISBN-10: 1-4051-6360-7 (hardback : acid-free paper) 1. Biomass energy. 2. Biomass conversion. 3. Plant cell walls. I. Himmel, Michael E.

TP339.B576 2008

662'.88—dc22

2007043423

A catalogue record for this book is available from the British Library.

Set in Minion 10/12pt by Aptara Inc., New Delhi, India

Printed and bound in Singapore by Fabulous Printers Pte Ltd

Contents

<i>Preface</i>	xiv
<i>Acknowledgments</i>	xv
<i>Contributors</i>	xvi
1 Our Challenge Is to Acquire Deeper Understanding of Biomass Recalcitrance and Conversion	1
<i>Michael E. Himmel and Stephen K. Picataggio</i>	
1.1 The modern lignocellulose biorefinery	1
1.2 Biomass recalcitrance to deconstruction	1
1.3 Plants evolved to resist microbial and enzymatic assault!	2
1.4 Are biomass-degrading enzymes working maximally?	2
1.5 Chemical pretreatments are still required to reveal cell wall cellulose	3
1.6 Fermenting cell wall sugars: the stage is set for systems/synthetic biology	4
References	5
2 The Biorefinery	7
<i>Thomas D. Foust, Kelly N. Ibsen, David C. Dayton, J. Richard Hess, and Kevin E. Kenney</i>	
2.1 Introduction	7
2.2 Phase III – lignocellulosic biorefineries	10
2.2.1 Feedstocks	12
2.2.2 Biochemical conversion	17
2.3 Thermochemical biorefinery	23
2.3.1 Introduction	23
2.3.2 R&D needs to achieve economic viability	26
2.4 Advanced biorefinery	28
2.4.1 Advanced, large-tonnage feedstock supply systems	28
2.4.2 Systems biology to improve biochemical processing	30

2.4.3	Selective thermal transformation to improve thermochemical processing	32
2.4.4	Technology integration, economies of scale, and evolutionary process optimization	34
	References	35
3	Anatomy and Ultrastructure of Maize Cell Walls: An Example of Energy Plants	38
	<i>Shi-You Ding and Michael E. Himmel</i>	
3.1	Introduction	38
3.2	Cell wall anatomy	38
3.2.1	Plant tissues	39
3.3	Cell wall biosynthesis and molecular structure	41
3.3.1	Biosynthesis	42
3.3.2	Cell wall lamellae	45
3.3.3	The microfibril and elementary fibril	46
3.3.4	The microfibril	47
3.3.5	Cellulose	48
3.3.6	Matrix polymers	49
3.4	Advanced approaches for characterizing cell wall structure	49
3.4.1	Atomic force microscopy	49
3.4.2	Biophotonics and nonlinear microscopy	50
3.4.3	Single molecule methods	50
3.4.4	Computer simulations	51
3.5	Summary	53
	Acknowledgment	55
	References	55
4	Chemistry and Molecular Organization of Plant Cell Walls	61
	<i>Philip J. Harris and Bruce A. Stone</i>	
4.1	Introduction	61
4.2	Chemistry of cell wall polymers	62
4.2.1	Chemistry of cell wall polysaccharides	62
4.2.2	Chemistry of cell wall proteins	70
4.3	Molecular associations between wall polymers	70
4.3.1	Non-covalent interactions between wall polymers	70
4.3.2	Covalent interactions between wall polymers	71
4.3.3	Covalent cross-linking between wall polymers prevents polysaccharide utilization	78
4.4	Molecular architecture of plant cell walls	79
4.4.1	Primary cell walls	79
4.4.2	Lignified secondary walls	81
4.5	Degradabilities of the walls of different cell types by enzymes	83
	References	85

5	Cell Wall Polysaccharide Synthesis	94
	<i>Debra Mohnen, Maor Bar-Peled, and Chris Somerville</i>	
5.1	Introduction	94
5.2	Cellulose	96
5.2.1	Enzymology	98
5.2.2	Cellulose deposition	100
5.2.3	Regulation of cellulose synthesis	101
5.3	Hemicellulose	104
5.3.1	Mannan	104
5.3.2	Xyloglucan	105
5.3.3	Xylan	108
5.3.4	Mixed linkage glucans	110
5.4	Pectins	110
5.4.1	Location of pectin synthesis	114
5.4.2	Pectin biosynthetic glycosyltransferases	115
5.4.3	Methyltransferases	119
5.4.4	Acetyltransferases	119
5.4.5	Other pectin modifying enzymes	119
5.4.6	Homogalacturonan synthesis	120
5.4.7	Xylogalacturonan synthesis	127
5.4.8	Apiogalacturonan synthesis	127
5.4.9	Synthesis of rhamnogalacturonan II (RG-II)	128
5.4.10	Rhamnogalacturonan I (RG-I) synthesis	130
5.5	The cell biology and compartmentalization of cell wall synthesis	136
5.6	Nucleotide sugars	137
5.6.1	Fermentation and nucleotide-sugars: a long history	140
5.6.2	Sugar kinase – pyrophosphorylase pathway to synthesize NDP-sugars	140
5.6.3	Direct production of NDP-sugars	140
5.6.4	NDP-sugar Interconversion Pathway	140
5.6.5	SLOPPY, a general UDP-sugar pyrophosphorylase	142
5.6.6	UDP- α -D-glucose (UDP-Glc)	143
5.6.7	ADP- α -D-glucose (ADP-Glc)	145
5.6.8	UDP- α -D-galactose (UDP-Gal)	146
5.6.9	UDP-L-rhamnose (UDP-Rha)	147
5.6.10	UDP- α -D-glucuronic acid (UDP-GlcA)	147
5.6.11	UDP- α -D-galacturonic acid (UDP-GalA)	150
5.6.12	UDP- α -D-xylose (UDP-Xyl)	151
5.6.13	UDP-D-apiose (UDP-Api)	151
5.6.14	UDP-L-arabinose pyranose (UDP-Ara)	152
5.6.15	UDP-arabinose furanose (UDP-Araf)	153
5.6.16	GDP- α -D-mannose (GDP-Man)	153
5.6.17	GDP- β -L-fucose (GDP-Fuc)	154
5.6.18	GDP- β -L-galactose (GDP-Gal), GDP- β -L-glucose gulose (GDP-Gul)	154

5.6.19	CMP- β -KDO (CMP-KDO)	154
5.6.20	Other enzymes involved in NDP-sugar metabolism	155
5.6.21	Future questions and directions	156
5.7	Perspectives	159
	Acknowledgments	159
	References	159
6	Structures of Plant Cell Wall Celluloses	188
	<i>Rajai H. Atalla, John W. Brady, James F. Matthews, Shi-You Ding, and Michael E. Himmel</i>	
6.1	Introduction	188
6.2	Background	189
6.3	Cellulose microfibrils	190
6.3.1	Molecular modeling	194
6.3.2	Raman spectra	200
6.3.3	Alternative patterns of aggregation	203
6.4	Alternative approaches to the problem of crystallinity	210
	References	210
7	Lignins: A Twenty-First Century Challenge	213
	<i>Laurence B. Davin, Ann M. Patten, Michaël Jourdes, and Norman G. Lewis</i>	
7.1	Lignin: molecular basis and role in plant adaptation to land	213
7.2	Lignin pathway evolution, deposition, and function in vascular anatomical development	218
7.2.1	Vascular plant diversification and lignification	218
7.2.2	Heartwood and reaction (compression/tension) wood tissues	223
7.3	Pioneers of monolignol biosynthesis, recent progress, and metabolic flux analyses	225
7.3.1	Phenylalanine formation	226
7.3.2	Metabolic flux analyses and transcriptional profiling in the monolignol pathway	227
7.3.3	Phenylalanine and tyrosine ammonia lyases	227
7.3.4	Cytochrome P-450s and hydroxycinnamoyl CoA:shikimate/quinic acid hydroxycinnamoyl transferases	228
7.3.5	4-Coumarate CoA ligases	229
7.3.6	Cinnamoyl CoA reductases and cinnamyl alcohol dehydrogenases	230
7.3.7	COMTs and CCOMTs	230
7.3.8	Proteins of unknown physiological/biochemical functions in monolignol metabolism, "CAD1" and "sinapyl alcohol dehydrogenase, SAD"	232
7.4	Recent developments: metabolic networks in the monolignol/lignin forming pathway (<i>Arabidopsis</i>) and (current) database annotations/limitations – opportunities and challenges	234

7.5	Inherent shortcomings in lignin analyses: a critical juncture and the urgent need	235
7.5.1	Lignin isolation procedures	236
7.5.2	Lignin subunit and lignin structural analyses by NMR spectroscopy	237
7.5.3	Quantification of lignin amounts, lignin degradation protocols, and synthetic dehydropolymerizates	239
7.6	Modulation of monolignol pathway and peroxidase enzymatic steps: predictable effects on the vascular apparatus and on limited substrate degeneracy during proposed lignin template polymerization	242
7.6.1	PAL, C4H, <i>p</i> C3H, HCT, and 4CL downregulation/mutation	243
7.6.2	CCR, CAD, F5H, and COMT downregulation/mutation, and the enigma of monolignol radical generation	254
7.6.3	Transcriptional control over secondary wall fiber formation: ramifications for lignification and vascular integrity	268
7.7	Native lignin macromolecular configuration	268
7.7.1	Early beginnings: the Freudenberg (random coupling) and the Forss (regular repeating unit) models for lignins	269
7.7.2	Further refinement of structural depictions of lignins (1970s to the present date): a reassessment	272
7.7.3	A new beginning: the need to fully define native lignin macromolecular configuration proper	274
7.8	Future outlook: remaining questions in lignin macromolecular assembly/configuration, proposed lignin template replication, and overall cell wall formation	285
	Acknowledgments	287
	References	287
8	Computational Approaches to Study Cellulose Hydrolysis	306
	<i>Michael F. Crowley and Ross C. Walker</i>	
8.1	Introduction	306
8.2	Molecular mechanics	307
8.2.1	The force field equation	307
8.2.2	Interatomic potentials	308
8.2.3	Non-bonded cutoffs and long range electrostatics	311
8.2.4	Molecular model types	312
8.3	Force fields	313
8.3.1	Carbohydrate force fields	314
8.3.2	Solvent models	314
8.4	Molecular dynamics	315
8.4.1	Dynamics methods	316
8.4.2	Finite difference methods	316
8.4.3	System size limitations	316
8.4.4	Quantum mechanics/molecular dynamics	317
8.5	Analysis methods	317

8.6	Enhanced sampling and free energy methods	319
8.6.1	Free energy methods	320
8.7	Studying cellulose hydrolysis	322
8.7.1	Work to date	322
8.7.2	Approaches to current questions about structure and hydrolysis	323
8.8	Performance and future of cellulose modeling	324
8.8.1	Current performance	324
8.9	Future possibilities	325
	Acknowledgments	326
	References	326
9	Mechanisms of Xylose and Xylo-oligomer Degradation During Acid Pretreatment	331
	<i>Xianghong Qian and Mark R. Nimlos</i>	
9.1	Background	331
9.2	Computational techniques	333
9.2.1	Molecular dynamics simulations	333
9.2.2	Static electronic structure theory	334
9.3	Xylose degradation reactions in vacuum	335
9.4	Effects of solvent water molecules	339
9.5	Xylobiose calculations	340
9.6	Experimental investigation of hydrolysis	344
9.7	The hydrolysis of xylobiose	345
9.8	The hydrolysis of xylan	346
9.9	Corn stover	347
9.10	Conclusions	348
9.11	Future studies	349
	Acknowledgment	349
	References	349
10	Enzymatic Depolymerization of Plant Cell Wall Hemicelluloses	352
	<i>Stephen R. Decker, Matti Siika-aho, and Liisa Viikari</i>	
10.1	Introduction	352
10.2	Hemicellulase types, activities, and specificities	355
10.3	Depolymerases	359
10.3.1	Xylanases	359
10.3.2	Mannanases	360
10.3.3	β -glucanases	361
10.3.4	Xyloglucanases	362
10.4	Debranching enzymes (accessory enzymes)	362
10.4.1	α -glucuronidase	363
10.4.2	α -arabinofuranosidase	363
10.4.3	α -D-galactosidase	363
10.4.4	Acetyl xylan esterase	363
10.4.5	Ferulic acid esterase	364

10.5	Hemicellulase activities for biomass feedstocks	364
10.5.1	Xylan	365
10.5.2	Galactoglucomannan and glucomannan	366
10.5.3	Arabinogalactan, xyloglucan, and β -glucan	367
10.6	Hydrolysis of solubilized hemicellulose	367
	Acknowledgment	368
	References	368
11	Aerobic Microbial Cellulase Systems	374
	<i>David B. Wilson</i>	
11.1	Introduction	374
11.2	Understanding cellulases	375
11.3	Diversity of cellulases	376
11.4	Cellulose-binding domains	379
11.5	Cellulase synergism	380
11.6	Cellulases from <i>Trichoderma reesei</i>	380
11.7	Other fungal cellulases	381
11.8	Cellulolytic aerobic bacteria	382
11.9	Outlook	386
	References	386
12	Cellulase Systems of Anaerobic Microorganisms from the Rumen and Large Intestine	393
	<i>Harry J. Flint</i>	
12.1	Introduction	393
12.2	Cellulolytic and hemicellulolytic bacteria from the rumen	394
12.2.1	<i>Ruminococcus flavefaciens</i>	394
12.2.2	Other Clostridium-related anaerobic bacteria	396
12.3	Plant cell wall breakdown by eukaryotic microorganisms	398
12.3.1	Rumen fungi	398
12.3.2	Rumen protozoa	398
12.4	Information from metagenomics	399
12.5	The large intestine	400
12.6	Conclusions	400
	Acknowledgment	401
	References	401
13	The Cellulosome: A Natural Bacterial Strategy to Combat Biomass Recalcitrance	407
	<i>Edward A. Bayer, Bernard Henrissat, and Raphael Lamed</i>	
13.1	Introduction	407
13.2	The cellulosome concept	408
13.3	Cellulosomal carbohydrate-active enzymes	410
13.4	The cellulosome–cellulose interaction	415
13.5	Cell-surface disposition of cellulosomes	417

13.6	Cellulosome assault on recalcitrant cellulose substrates	418
13.7	Degradation of cellulose by the <i>C. thermocellum</i> cellulosome	420
13.8	The cellulosome rationale	423
	Acknowledgments	426
	References	426
14	Pretreatments for Enhanced Digestibility of Feedstocks	436
	<i>David K. Johnson and Richard T. Elander</i>	
14.1	Introduction	436
14.2	Enzyme usage and enzyme-type considerations for pretreated biomass	437
14.3	Desired properties of pretreatment processes	437
14.4	Physicochemical properties of pretreated biomass believed to affect cellulose digestibility	438
14.5	Pretreatment approaches	439
14.5.1	Physical pretreatments	440
14.5.2	Rapid decompression pretreatments	440
14.5.3	Autohydrolysis pretreatments	442
14.5.4	Acidic pretreatments	443
14.5.5	Alkaline pretreatments	444
14.5.6	Solvent pretreatments	445
14.5.7	Supercritical fluid pretreatments	446
14.5.8	Oxidative pretreatments	446
14.5.9	Biological pretreatment	447
14.6	Future prospects	447
	Acknowledgment	449
	References	449
15	Understanding the Biomass Decay Community	454
	<i>William S. Adney, Daniel van der Lelie, Alison M. Berry, and Michael E. Himmel</i>	
15.1	Introduction	454
15.2	Defining biomass decay communities	456
15.2.1	Fungi identified with plant biomass	457
15.2.2	Bacteria identified with plant biomass	459
15.3	Interactions between saprophytic fungi and bacteria	463
15.4	Characterization of microbial communities that degrade biomass	464
15.4.1	Biochemical approaches to define biomass degrading communities	465
15.4.2	Molecular approaches for defining biomass-degrading communities	466
15.4.3	Microarray methods suitable for biomass sampling	470
15.5	Conclusions	472
	Acknowledgment	472
	References	473

16 New Generation Biomass Conversion: Consolidated Bioprocessing	480
<i>Y.-H. Percival Zhang and Lee R. Lynd</i>	
16.1 Introduction	480
16.2 Consolidated bioprocessing	481
16.3 CBP advances	483
16.3.1 Native cellulolytic microorganisms	483
16.3.2 Recombinant cellulolytic strategy	488
16.4 Future directions	489
Acknowledgment	490
References	490
 <i>Index</i>	 495
<i>Color plate section after page 238</i>	

Preface

The chapters in this book describe the state of the art, as well as promising new approaches, to overcome the critical science and engineering barriers to enabling modern biorefineries. We have assembled chapters that focus on topics extending from the highest levels of biorefinery design and biomass life-cycle analysis, to detailed aspects of plant cell wall structure, chemical treatments, enzymatic hydrolysis, and product fermentation options. Such compendia are often mere signposts in time. However, we hope that our unique assembly of carefully integrated topics, presented with reviews of background science, will remain relevant for those working in the biomass conversion field.

Acknowledgments

The authors wish to thank Todd Vinzant for the scanning electron micrograph image used on the cover of this book.

The U.S. Department of Energy Office of the Biomass Program is acknowledged for supporting the compilation of this book as well as its almost three decades of support for the biomass conversion sciences. Without this visionary programmatic perspective, much of the work reported here would not have been possible.

Contributors

William S. Adney	National Renewable Energy Laboratory, Chemical and Biosciences Center, Golden, CO 80401, USA
Rajai H. Atalla	Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI 53705 USA
Maor Bar-Peled	Complex Carbohydrate Research Center, Department of Plant Biology, BioEnergy Science Center, The University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA
Edward A. Bayer	Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel
Alison M. Berry	Department of Environmental Horticulture, University of California, One Shields Avenue, Davis, CA 95616, USA
John W. Brady	Department of Food Science, 101 Stocking Hall, Cornell University, Ithaca, NY 14853-7201, USA
Michael F. Crowley	Department of Molecular Biology, TPC6, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
Laurence B. Davin	Washington State University, Institute of Biological Chemistry, Pullman, WA 99164-6340, USA
David C. Dayton	National Renewable Energy Laboratory, National Bioenergy Center, Golden, CO 80401, USA
Stephen R. Decker	National Renewable Energy Laboratory, Chemical and Biosciences Center, Golden, CO 80401, USA
Shi-You Ding	National Renewable Energy Laboratory, Chemical and Biosciences Center, Golden, CO 80401, USA
Richard T. Elander	National Renewable Energy Laboratory, National Bioenergy Center, Golden, CO 80401, USA

Harry J. Flint	Microbial Ecology Group, Rowett Research Institute, Greenburn Road, Bucksburn, Aberdeen AB21 9SB, UK
Thomas D. Foust	National Renewable Energy Laboratory, National Bioenergy Center, Golden, CO 80401, USA
Philip J. Harris	School of Biological Sciences, The University of Auckland, Auckland, New Zealand
Bernard Henrissat	Architecture et Fonction des Macromolécules Biologiques, CNRS, Universités Aix-Marseille I & II, 13288 Marseille Cedex 9, France
J. Richard Hess	INL Bioenergy Program Technology, Idaho National Laboratory, PO Box 1625, Idaho Falls, ID 83415 USA
Michael E. Himmel	National Renewable Energy Laboratory, Chemical and Biosciences Center, Golden, CO 80401, USA
Kelly N. Ibsen	National Renewable Energy Laboratory, National Bioenergy Center, Golden, CO 80401, USA
David K. Johnson	National Renewable Energy Laboratory, Chemical and Biosciences Center, Golden, CO 80401, USA
Michaël Jourdes	Washington State University, Institute of Biological Chemistry, Pullman, WA 99164-6340, USA
Kevin E. Kenney	INL Bioenergy Program Technology, Idaho National Laboratory, PO Box 1625, Idaho Falls, ID 83415, USA
Raphael Lamed	Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv 69978, Israel
Norman G. Lewis	Washington State University, Institute of Biological Chemistry, Pullman, WA 99164-6340, USA
Lee R. Lynd	Thayer School of Engineering, Dartmouth College, 8000 Cummings Hall, Hanover, NH 03755-8000, USA
James F. Matthews	Department of Food Science, 101 Stocking Hall, Cornell University, Ithaca, NY 14853-7201, USA
Debra Mohnen	Complex Carbohydrate Research Center, Department of Biochemistry and Molecular Biology, BioEnergy Science Center, The University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA
Mark R. Nimlos	National Renewable Energy Laboratory, National Bioenergy Center, Golden, CO 80401, USA
Ann M. Patten	Washington State University, Institute of Biological Chemistry, Pullman, WA 99164-6340, USA

Stephen K. Picataggio	Synthetic Genomics, Inc., 11149 North Torrey Pines Road, La Jolla, CA 92037, USA
Xianghong Qian	Department of Mechanical Engineering, Colorado State University, 1374 Campus Delivery, Fort Collins, CO 80523-1374, USA
Matti Siika-aho	VTT Biotechnology, Tietotie 2, FI-02044, Espoo, Finland
Chris Somerville	Energy Biosciences Institute, University of California, Berkeley, 130 Calvin Laboratory, Berkeley CA 94720-5230, USA
Bruce A. Stone	Department of Biochemistry, La Trobe University, Melbourne 3086, Australia
Daniel van der Lelie	Biology Department, 463, Brookhaven National Laboratory, Upton, NY 11973-5000, USA
Liisa Viikari	VTT Biotechnology and Food Research, Tietotie 2, FI-02044, Espoo, Finland
Ross C. Walker	San Diego Supercomputer Center, UC San Diego, MC 0505, 9500 Gilman Drive, La Jolla, CA 92093-0505, USA
David B. Wilson	Department of Molecular Biology and Genetics, 458 Biotechnology Building, Cornell University, Ithaca, NY 14853, USA
Y.-H. Percival Zhang	Biological Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

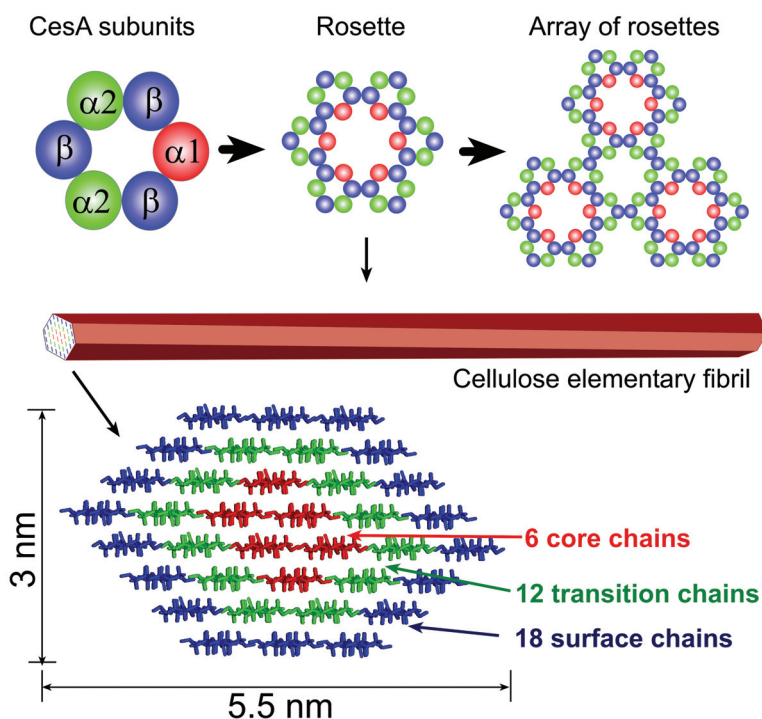


Plate 1 Model of plant cell wall cellulose elementary fibril and its synthesis. In this model, at least three types of cellulose synthases (CesA subunits, $\alpha 1$, $\alpha 2$, and β) are needed to spontaneously assemble the rosettes that composed of 6×6 CesA enzymes synthesizing 36-chain cellulose elementary fibril. The rosettes may also form arrays in the cell membrane, in this case, a number of rosettes synthesize a bundle of elementary fibril, the macrofibril. The estimated dimensions of elementary fibril are 3×5.5 nm that agrees with direct measurement using atomic force microscopy (see also Figure 3.9). The depiction of the glucan chains is based generally on an X-ray structure of cellulose I β . It has been proposed that the cellulose elementary fibril may contain three groups of glucan chains: in group C1 (red) there are 6 crystalline chains; in group C2 (green) there are 12 sub-crystalline chains with a small degree of disorder; and in group C3 (blue) there are 18 surface chains that are sub-crystalline with a large degree of disorder. (Modified from Ding *et al.*, 2006; Himmel *et al.*, 2007.) (See Figure 3.4.)

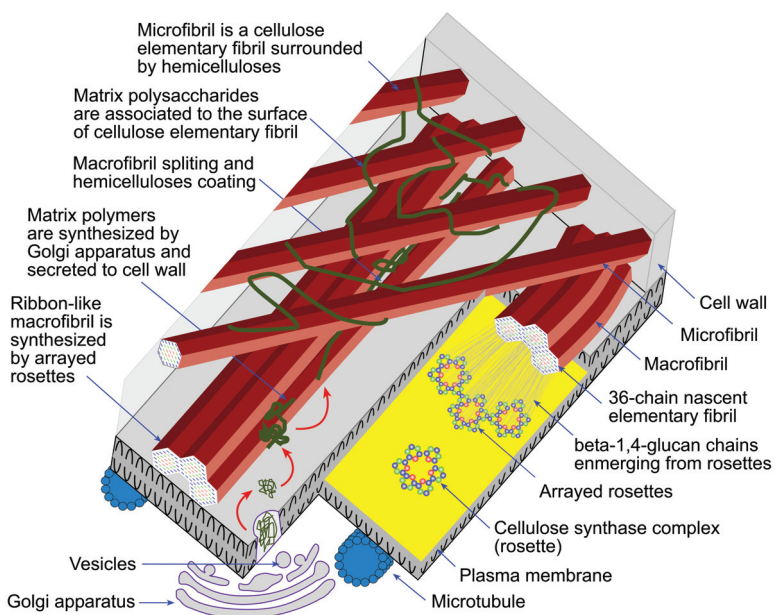


Plate 2 Schematic diagram of plant cell wall synthesis. (See Figure 3.5.)

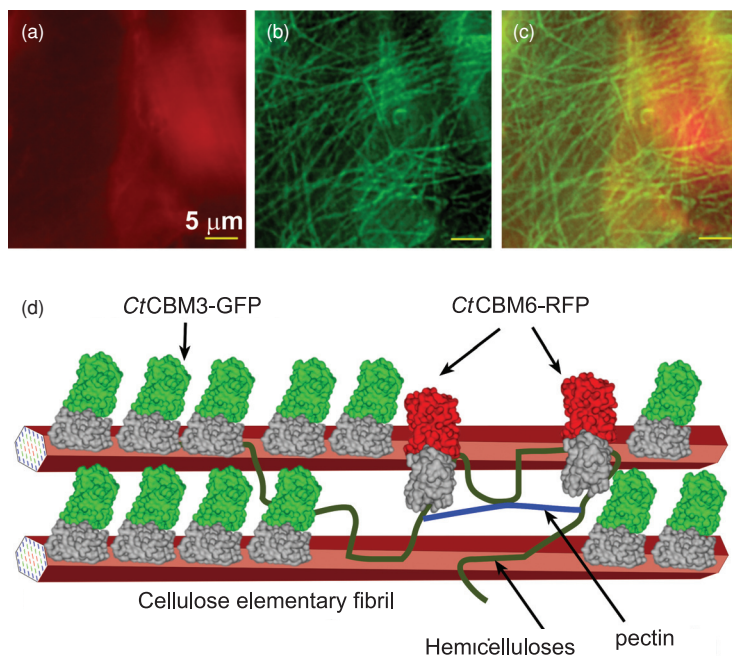


Plate 3 Total internal reflection fluorescence micrograph of fresh maize parenchyma cell wall labeled by CtCBM3-GFP and CtCBM6-RFP. (Modified from Ding *et al.*, 2006.) (See Figure 3.10.)

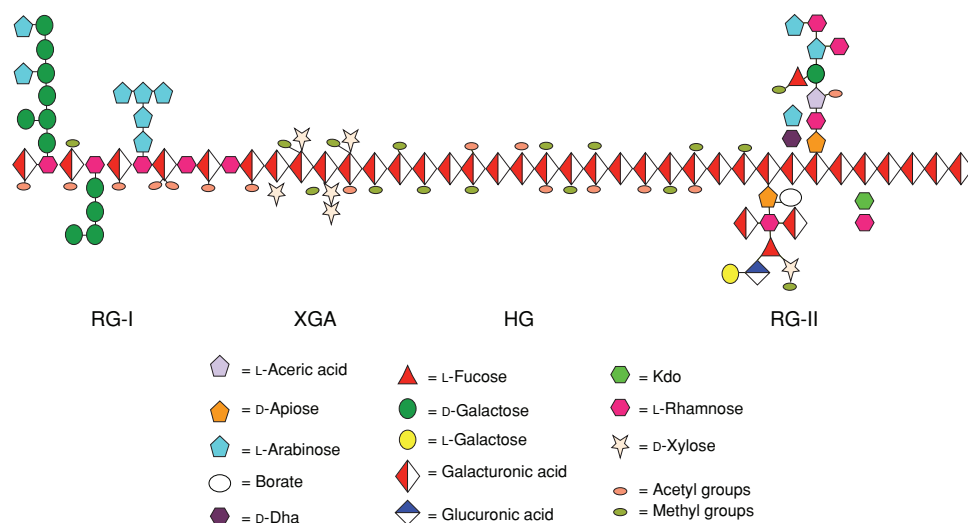


Plate 4 Schematic structure of pectin showing the three main pectic polysaccharides homogalacturonan (HG), rhamnogalacturonan I (RG-I), and rhamnogalacturonan II (RG-II) linked to each other. A region of substituted galacturonan known as xylogalacturonan is also shown (XGA). The representative pectin structure shown is not quantitatively accurate, HG should be increased 12.5-fold and RG-I increased 2.5-fold to approximate the amounts of these polysaccharides relative to each other plant walls. The monosaccharide symbols used are either from the Symbol and Text Nomenclature for Representation of Glycan Structure. Nomenclature Committee Consortium for Functional Glycomics (<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>) or from D. Mohnen. (The figure is modified from <http://www.uk.plbio.kvl.dk/plbio/cellwall.htm>.) (See Figure 5.3.)

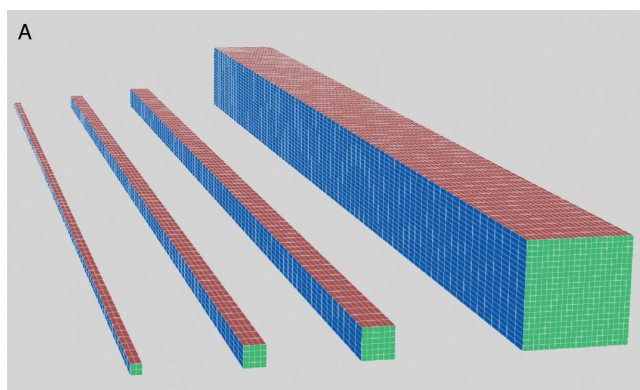


Plate 5 Geometric representation of nanofibrils that are 2 by 2, 4 by 4, 6 by 6, and 20 by 20 nm in cross section. Panel A illustrates nanofibrils in translational symmetry in three directions. (See Figure 6.3a.)

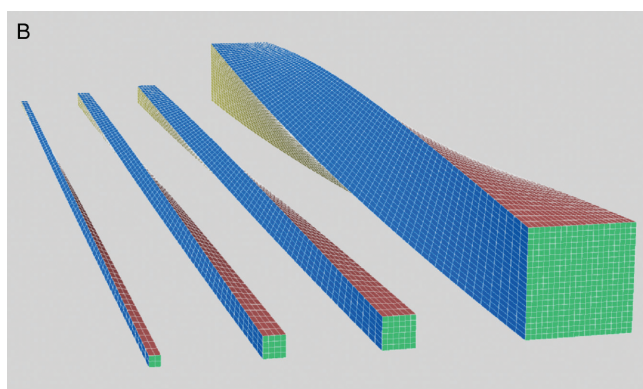


Plate 6 Panel B shows a long period of 1200 nm as shown in Figure 6.4. (See Figure 6.3b.)

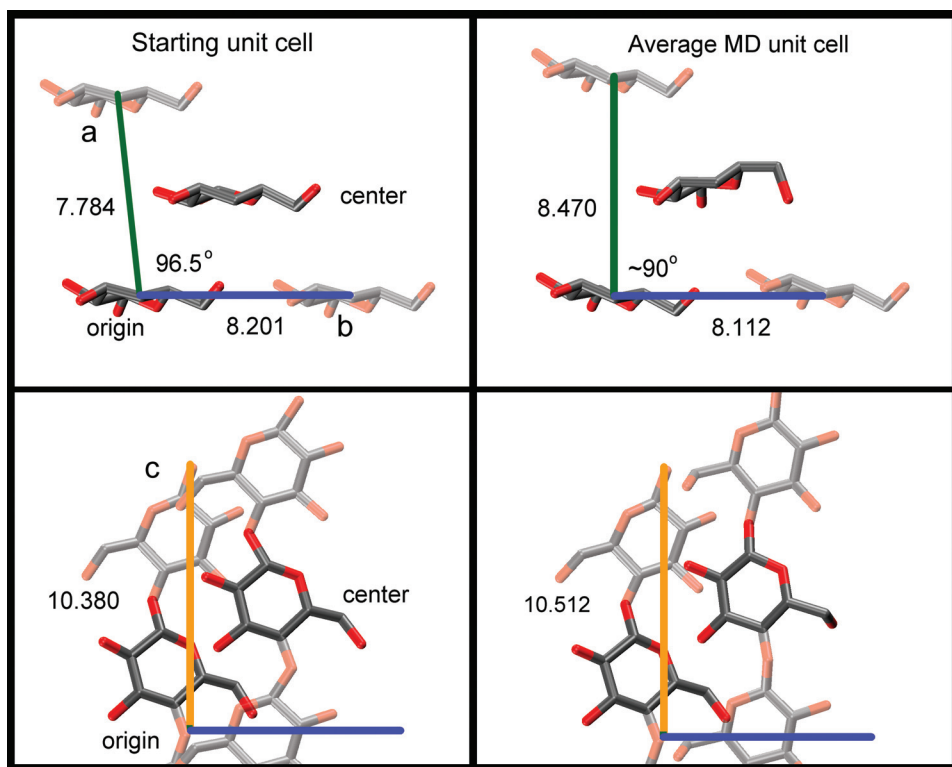


Plate 7 Left: the cellulose I_β crystal unit cell determined by fiber diffraction; right: the trajectory-averaged unit cell for the simulation of the diagonal crystal. Hydrogen atoms are omitted for clarity, and positions obtained by symmetry operations are transparent. (See Figure 6.4.)

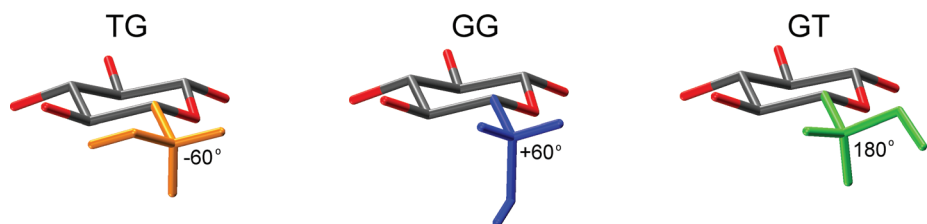


Plate 8 Nomenclature of primary alcohol conformation. The dihedral angle measured by C4–C5–C6–O6 is shown. (See Figure 6.5.)

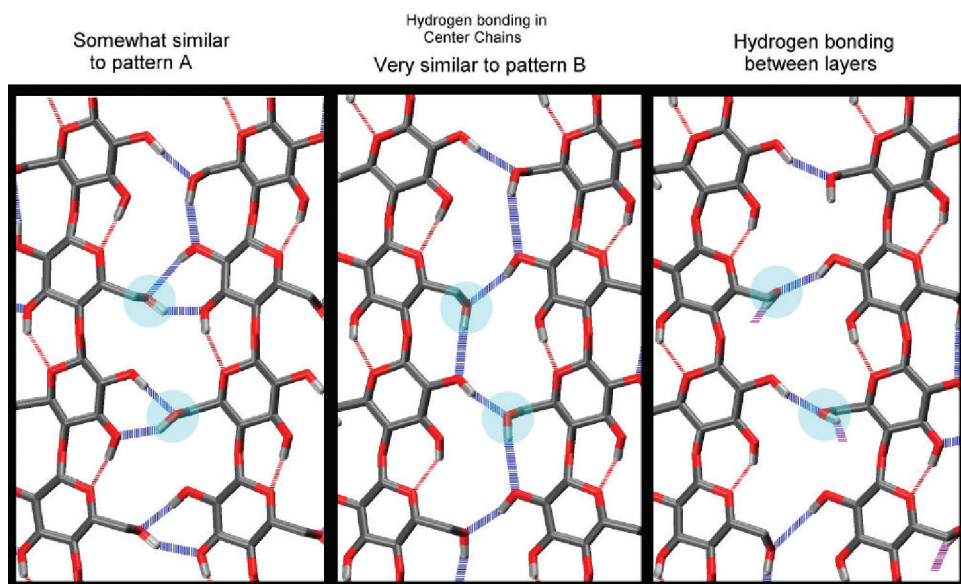


Plate 9 Single frames from the center chain layers illustrating three different hydrogen bond patterns. Left: the pattern is similar to the predominant pattern from the crystal structure, but the rotation to GG makes the HO2–O6 hydrogen bond across the glycosidic linkage impossible; center: the hydrogen bond pattern is very similar to the less occupied pattern from the crystal structure; right: hydrogen bonds from HO6 in a center chain to O2 in an origin layer chain, which is not shown for clarity. (See Figure 6.6.)

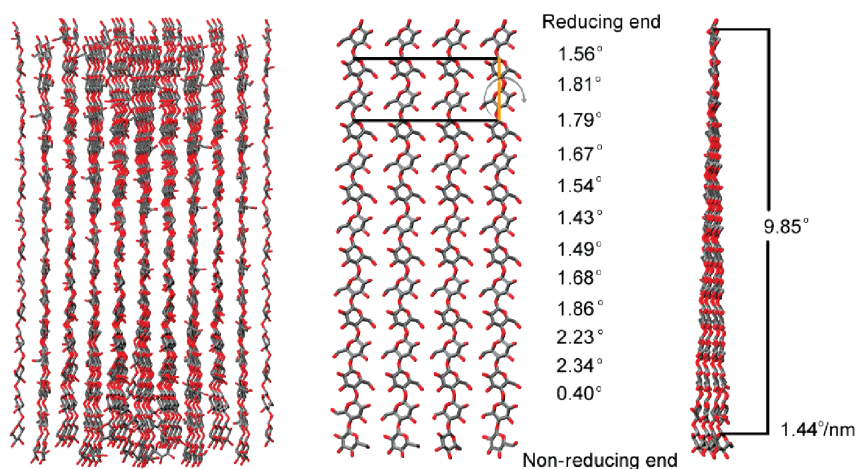


Plate 10 Views of the central portion of one center chain and one origin chain from the middle of the diagonal crystal, illustrating the inter-plane hydrogen bonds which can occur after the center chain primary alcohol groups rotate to the GG conformation. Hydrogen bonds between layers are indicated with dashed lines. (See Figure .6.7)

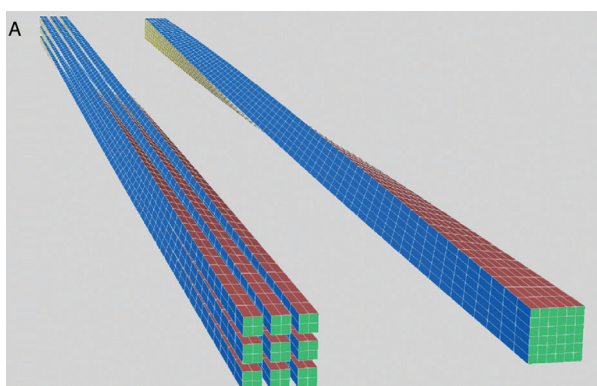


Plate 11 Effects of packing. Panel A compares the helical form of a 6 by 6 nm nanofibril with that of nine 2 by 2 helical nanofibrils packed as close as possible without surface intersections. (See Figure 6.11a.)

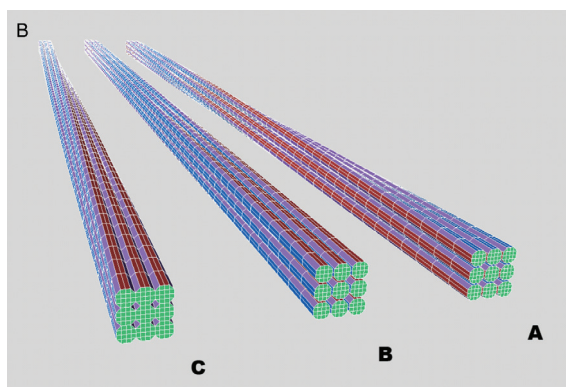


Plate 12 Panel B shows helical nanofibrils assembled in different patterns. (See Figure 6.11b.)

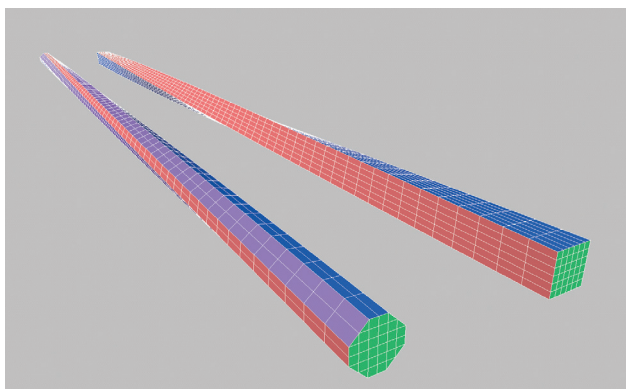


Plate 13 Aggregation into a larger nanofibril. (See Figure 6.12.)

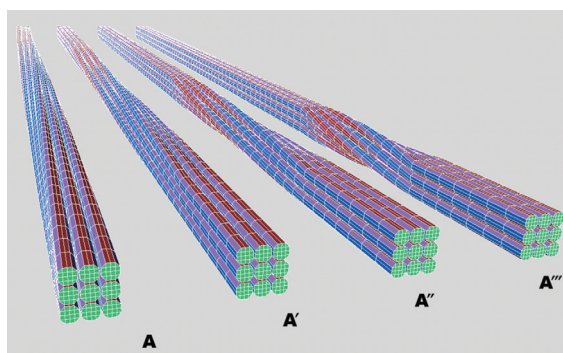


Plate 14 The secondary aggregate A is identical to the one designated as A on the right side of panel B in Figure 6.11, thus it has the primary aggregates twisted to match a 1200-nm period helix and then the secondary aggregate is subjected to a similar helical deformation; this is the uniform secondary aggregate proposed as the most efficient load-bearing structure. In the progression toward A', A'', and A''', different levels of further aggregation are depicted such that portions of the primary aggregates are made parallel and thus more closely approximate a lattice. In A', it is 30%, then 65% in A'', and finally 85% in A'''. (See Figure 6.14.)

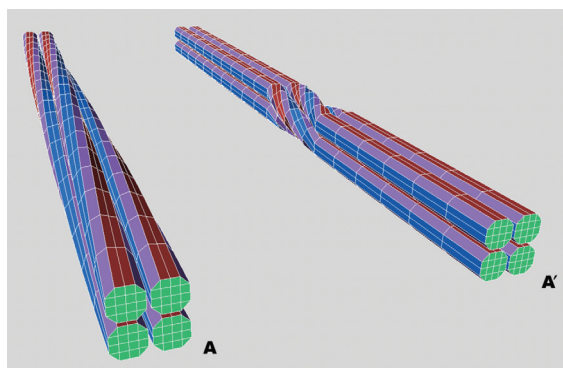


Plate 15 Representations of the patterns of aggregation in a fibril 4 nm in diameter, consisting of four elementary fibrils. In A, each of the elementary fibrils has a helical twist with long period of 300 nm and the collective aggregate also has a helical twist with a long period of 300 nm. In A', 85% of the length has been forced to align in a parallel pattern resulting in what can appear to be nanocrystalline domains. (See Figure 6.15.)

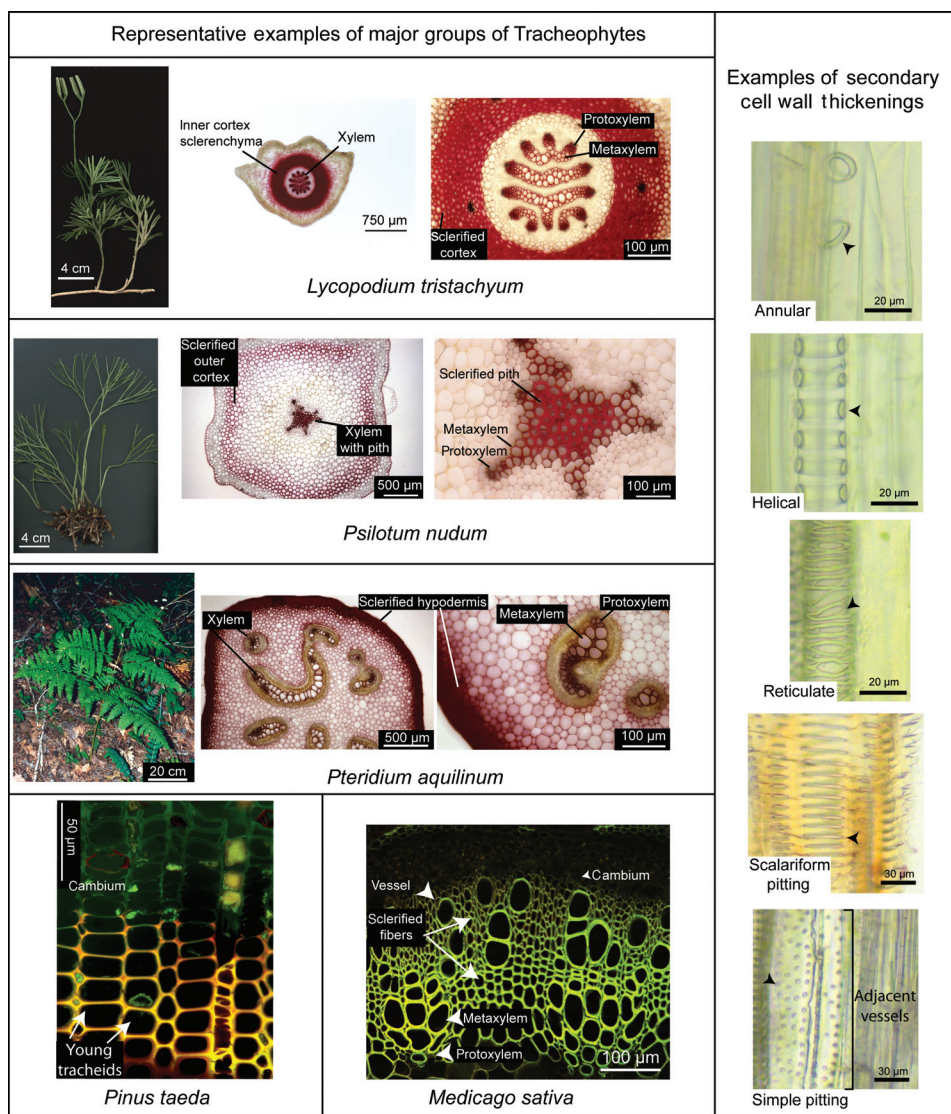


Plate 16 Lignified vascular and sclerenchyma anatomical structures in extant plant lineages. Proto- and metaxylem of the major plant lineages: Lycopodiophyta (e.g., *Lycopodium tristachyum*), Psilophyta (e.g., *Psilotum nudum*), Filicophyta (e.g., *Pteridium aquilinum*), gymnosperms (e.g., *Pinus taeda*), and angiosperms (e.g., *Medicago sativa*) have very similar annular and helical secondary cell wall thickenings. Some members of each group, but especially the Filicophyta, have reticulate tracheid cell wall thickenings, a more complex form than the helical form. Secondary cell wall thickenings with scalariform pitting may be found in all plant lineages and represent an intermediate form to the simple pitting found in higher plants (i.e., gymnosperms and angiosperms). Brightfield microscopy images of *L. tristachyum*, *P. nudum*, and *P. aquilinum* were taken of hand-cut fresh sections stained with phloroglucinol-HCl to reveal patterns of phenolic deposition (red) in the xylem and sclerified fiber cells. Epifluorescent confocal images of *P. taeda* and *M. sativa* were made using cryosections of fresh tissue stained with a combination of acridine orange and ethidium bromide to reveal patterns of lignification. Examples of secondary cell wall thickenings were recorded using brightfield microscopy and hand-cut longitudinal sections of fresh unstained xylem from *Equisetum telmateia* (a member of Equisetophyta for annular and helical examples), *Ophioglossum reticulatum* (a member of the Filicopsida, for the reticulate example), *P. aquilinum* (another member of the Filicopsida, for the scalariform-pitted example), and *M. sativa* (an angiosperm, for the example of vessel secondary cell wall thickenings with simple pitting). (See Figure 7.4.)

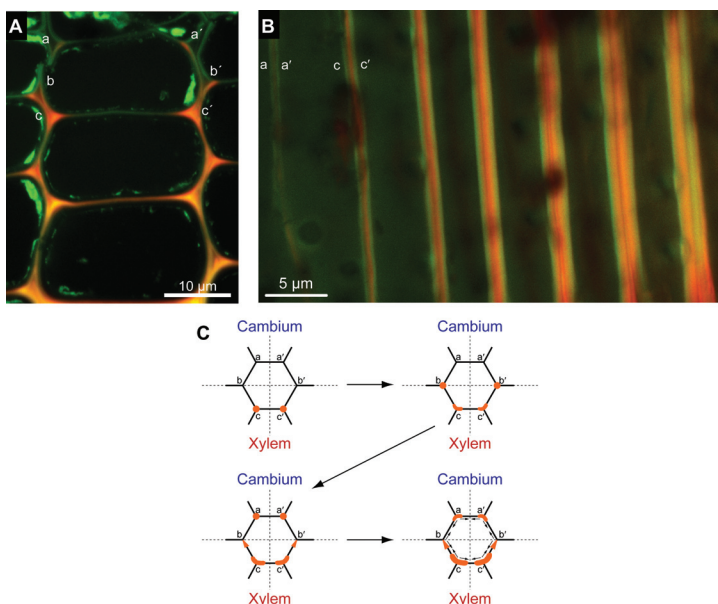


Plate 17 Lignification and cell wall development in loblolly pine cambial tissue (A and B) show deposition of lignin in both transverse and longitudinal sections (Y. Nakazawa, unpublished). In Fig. 5A, lignin is differentially initiated at sites a,a', b,b' and c,c'. In Fig. 5B, the particular cell has initiated lignification at cell corners c,c', but not yet at a,a'. Lignification is initiated at 3 points, c,c', b,b' then a,a' in a symmetrical but differential manner. [Cellulose/hemicellulose regions are visualized in green, and lignin is orange/yellow.] (C) Idealized depiction of differential lignin deposition in one tracheary cell type. → in A and • in C = lignin initiation sites. A dual-stain method using acridine orange (AO) and ethidium bromide (EB) was used to visualize lignin as a red-orange epifluorescence. The merged confocal images were individually recorded using detection filters specific to 488 nm (Emission) and 522 nm (Excitation) for AO and 568 nm (Ex.) and 598 nm (Em.) for EB. (See Figure 7.5.)

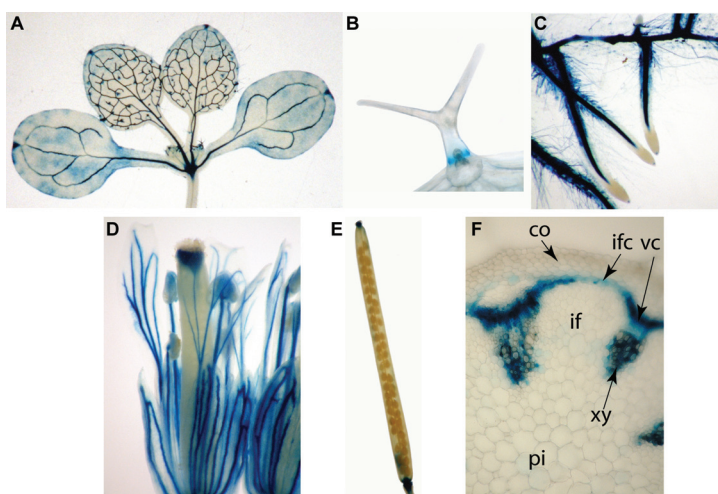


Plate 18 Cinnamyl alcohol dehydrogenase gene expression in the vascular apparatus of *Arabidopsis thaliana* (55). Selected examples of GUS-visualized expression of AtCAD5 (A, C–F) and AtCAD4 (B). (A) In vascular apparatus, including hydathodes of 2-week-old leaf tissue; (B) at the base of trichomes; (C) in primary and secondary roots; (D) in sepal/petal veins, style, anthers, and stamen filaments of the flower; (E) in the abscission and style regions of the silique; and (F) in vascular cambium, interfascicular cambium, and the developing xylem of the inflorescence stem. (Reprinted from Phytochemistry, vol. 68, Kim, S.-J., Kim, K.-W., Cho, M.-H., Franceschi, V.R., Davin, L.B. & Lewis, N.G., Expression of cinnamyl alcohol dehydrogenases and their putative homologues during *Arabidopsis thaliana* growth and development: Lessons for database annotations? pp. 1957–1974, Copyright 2007, with permission from Elsevier.) (See Figure 7.6.)

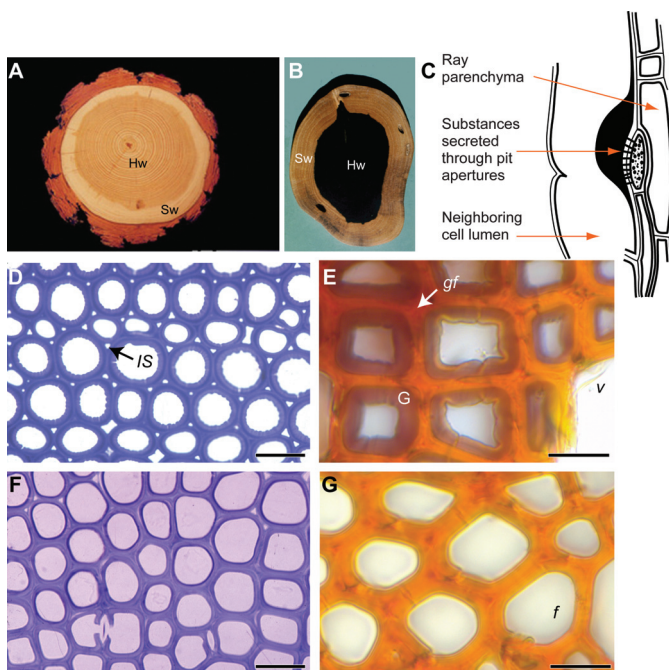


Plate 19 Woody cross sections showing heartwood deposition and reaction wood tissues. Heartwood in (A) tamarack (*Larix laricina*) and (B) ebony (*Diospyros ebenum*), as well as (C) secretion of heartwood constituents by ray parenchyma cells into lumen of neighboring cells; this appears to occur through pit apertures (73). Light micrograph cross section of compression (D) and "normal" (F) wood in Douglas fir (*Pseudotsuga menziesii*) (75) and of tension (E) and "normal" (G) wood in black cottonwood (*Populus balsamifera* ssp. *trichocarpa*) (72). Bar: 20 μ m (D, F) and 10 μ m (E, G). Abbreviations: *f*, fiber; G, G-layer; Hw, heartwood; IS, intercellular space; *gf*, gelatinous fiber; Sw, sapwood; *v*, vessel. [Reprinted from (A) *Phytochemistry*, vol. 57, Kwon, M., Bedgar, D.L., Piastuch, W., Davin, L.B. & Lewis, N.G., Induced compression wood formation in Douglas fir (*Pseudotsuga menziesii*) in microgravity, pp. 847–857, Copyright 2001, with permission from Elsevier. (B) *Current Opinion in Plant Biology*, vol. 2, Lewis, N.G., A 20th century roller coaster ride: A short account of lignification, pp. 153–162, Copyright 1999, with permission from Elsevier. (C) ACS Symposium Series, vol. 697, Gang, D.R., Fujita, M., Davin, L.B. & Lewis, N.G., The "abnormal lignins": Mapping heartwood formation through the lignan biosynthetic pathway, pp. 389–421, Copyright 1998, with permission from American Chemical Society. (E and F) *The American Journal of Botany*, vol. 94, Patten, A.M., Jourdes, M., Brown, E.E., Laborie, M.-P., Davin, L.B. & Lewis, N.G., Reaction tissue formation and stem tensile modulus properties in wild type and *p*-coumarate-3-hydroxylase downregulated lines of alfalfa, *Medicago sativa* (Fabaceae), pp. 912–925, Copyright 2007, with permission from the Botanical Society of America.] (See Figure 7.7.)

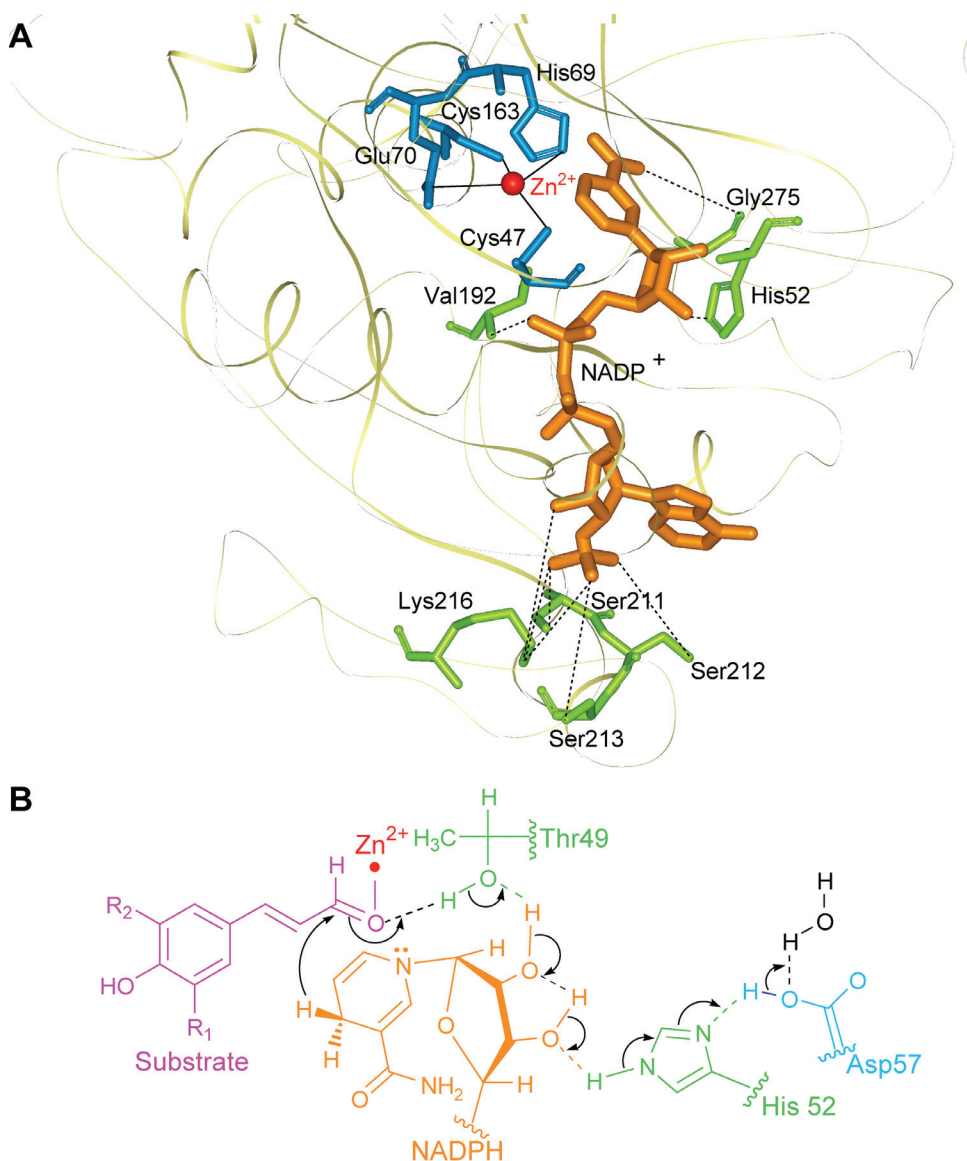


Plate 20 (A) Structure of the substrate-binding pocket of $NADP^+$ binary form of AtCAD5 showing the catalytic Zn^{2+} ion (red sphere) tetrahedrally coordinated by Cys47, His69, Cys163 and Glu70 (blue). The $NADP^+$ molecule (orange) is held by Val192, Ser211, Ser212, Ser213, Lys216 and Gly275 (green) (133). [Possible hydrogen bonds are shown as black dotted lines.] (B) Proposed proton shuttling mechanism during the reduction process in the active site of the AtCAD5. Solid arrows indicate the movement of two electrons among the functional groups during substrate reduction (133). The possible hydrogen bonds involved are shown with dotted lines. (Reprinted from Organic and Biomolecular Chemistry, vol. 4, Youn, B., Camacho, R., Moinuddin, S.G.A., Lee, C., Davin, L.B., Lewis, N.G. & Kang, C., Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4, pp. 1687–1697, Copyright 2006, with permission from The Royal Society of Chemistry.) (See Figure 7.9.)

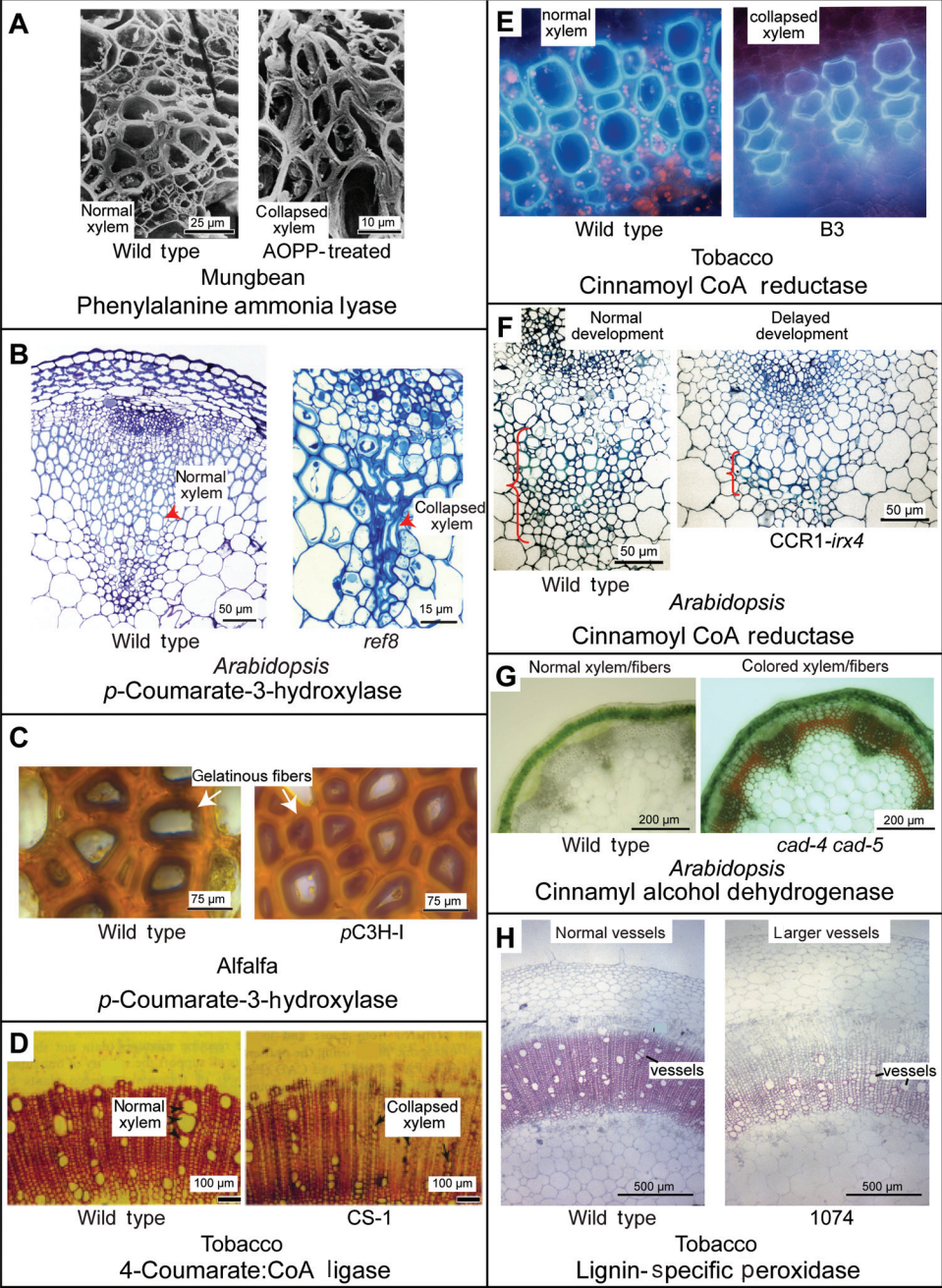


Plate 21 Examples of effects on vascular anatomy of inhibiting/mutating/downregulating various enzymes and/or genes in the phenylpropanoid pathway (see Figure 7.12 for details).

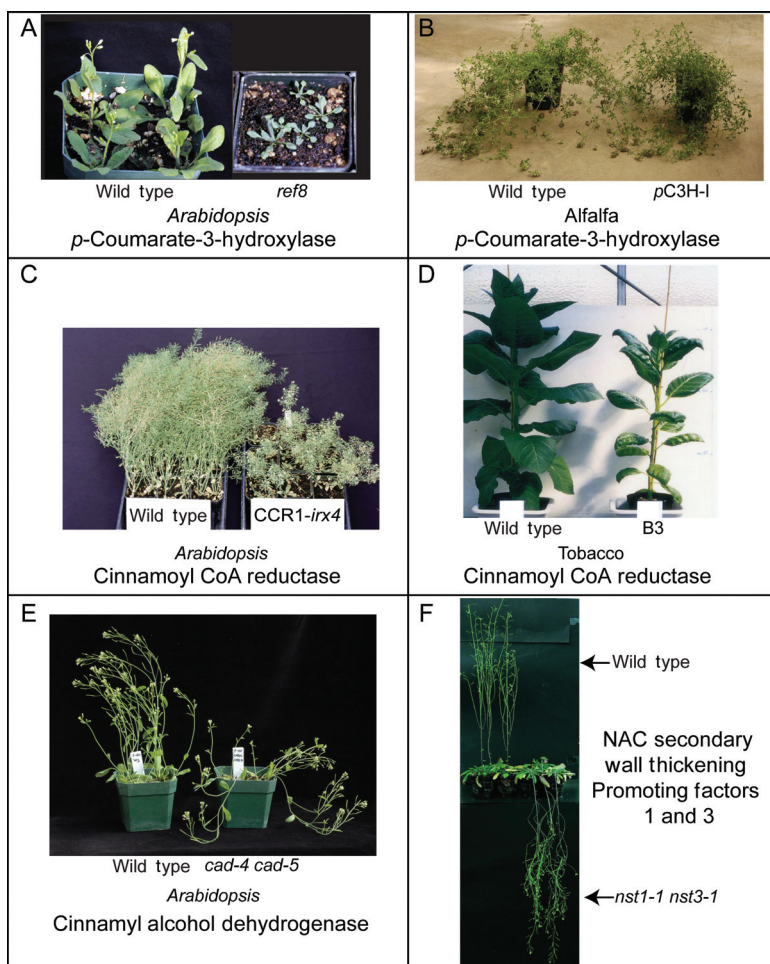


Plate 22 Gross phenotypical changes/effects of either mutating or downregulating various phenylpropanoid genes, as well as vascular related transcription factors. **(A)** *pC3H* mutant with various pleiotropic effects, resulted in a severely dwarfed line (*ref8*) (213, 215) (shown 23 days post seed-sown). **(B)** *pC3H* downregulation in alfalfa (*Medicago sativa*) resulted in a phenotype (*pC3H-I*) very similar to wild type without visible pleiotropic effects (shown 4 weeks post-cut-back). **(C)** Dwarfed *Arabidopsis* *CCR irx4* mutant with pleiotropic effects (shown 6 weeks post seed-sown) (131). **(D)** *CCR* downregulation in tobacco, with stunted morphology relative to wild type (232). **(E)** *CAD* double mutation in *Arabidopsis* (*AtCAD4/5*, *cad-4*, *cad-5*) resulted in a limp to prostrate stem phenotype (shown after 4 weeks growth) (71). **(F)** Double T-DNA tagged *Arabidopsis nst1 nst3* line (199). [Reprinted from **(B)** The American Journal of Botany, vol. 94, Patten, A.M., Jourdes, M., Brown, E.E., Laborie, M.-P., Davin, L.B. & Lewis, N.G., Reaction tissue formation and stem tensile modulus properties in wild type and *p*-coumarate-3-hydroxylase downregulated lines of alfalfa, *Medicago sativa* (Fabaceae), pp. 912–925, Copyright 2007, with permission from the Botanical Society of America. **(D)** Plant Journal, vol. 13, Piquemal, J., Lapierre, C., Myton, K., O'Connell, A., Schuch, W., Grima-Pettenati, J. & Boudet, A.-M., Downregulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants, pp. 71–83, Copyright 1998, with permission from Blackwell; and **(F)** The Plant Cell, vol. 19, Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K. & Ohme-Takagi, M., NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*, pp. 270–280, Copyright 2007, with permission from the American Society of Plant Biologists.] (See Figure 7.13.)

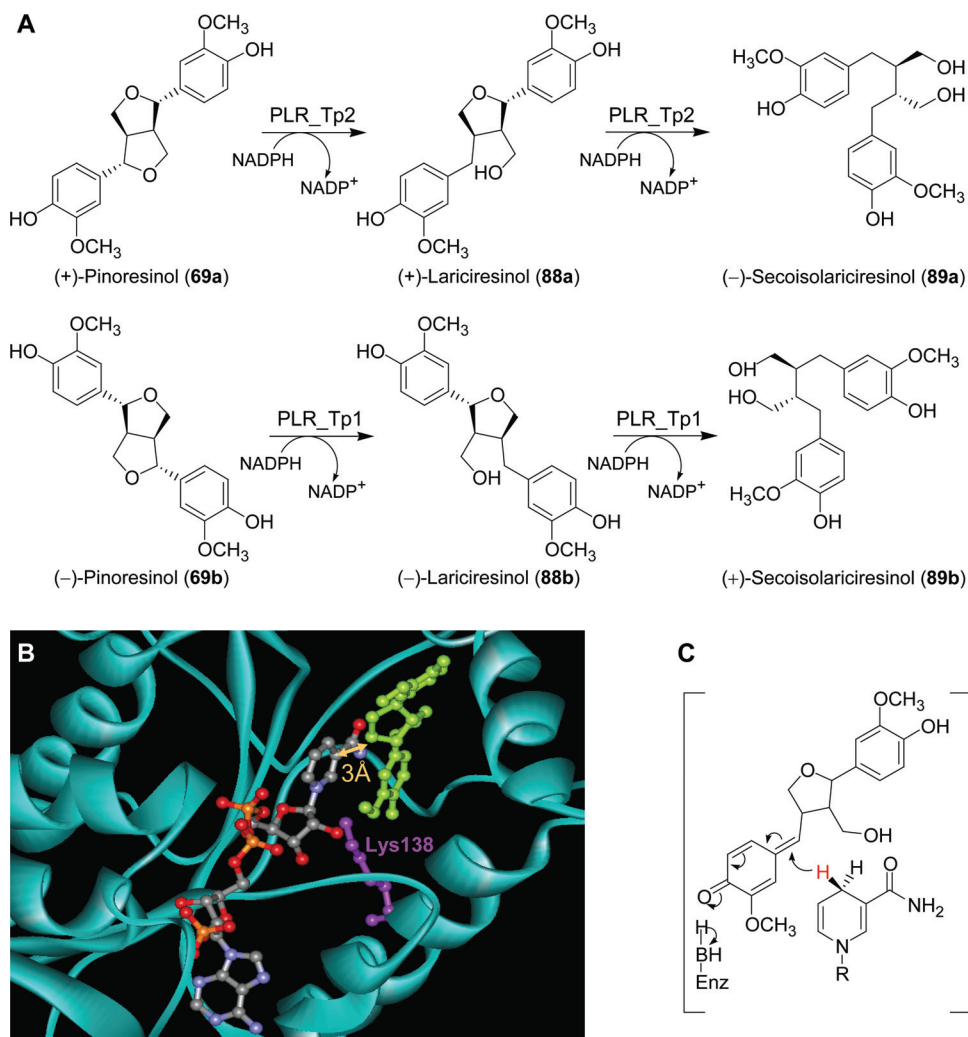


Plate 23 Pinoresinol-lariciresinol reductases (PLRs)/PLR homologues, PLR_Tp1 and PLR_Tp2, in the gymnosperm western red cedar: **(A)** Differing enantiospecificity differences of distinct PLR (homologue); **(B)** Partial crystal structure of PLR_Tp1 showing general catalytic base, Lys137, together with substrate (**69b**) and NADPH; and **(C)** Proposed sequential reduction to secoisolariciresinol (**89**) via intermediary quinone methide. (See Figure 7.22.)

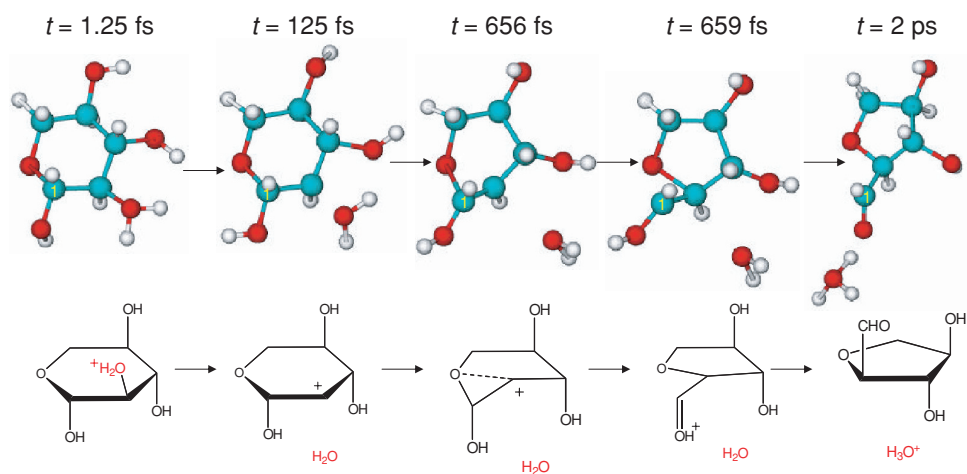


Plate 24 Results of CPMD simulation of xylose degradation after protonation of the hydroxyl group on O2. After 125 fs xylose is dehydrated and at approximately 659 fs the remaining carbocation rearranges to form the dehydrated furanyl form of xylose. This product will need to undergo two additional dehydrations to form furfural. (See Figure 9.2.)

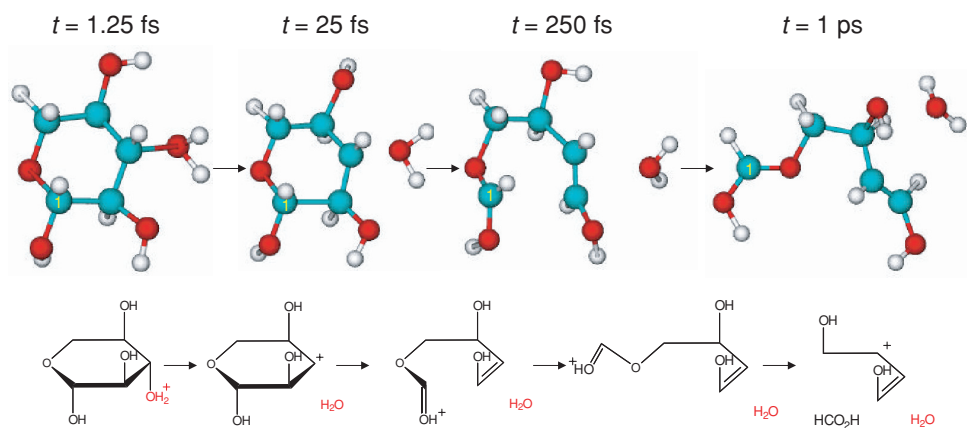


Plate 25 Results of CPMD simulation of xylose degradation after protonation of the hydroxyl group on O3. After 25 fs xylose is dehydrated and at approximately 250 fs the C1–C2 bond has broken. In this product the C1–O5 will eventually break to yield formic acid as shown in the Lewis structure at the end. (See Figure 9.3.)

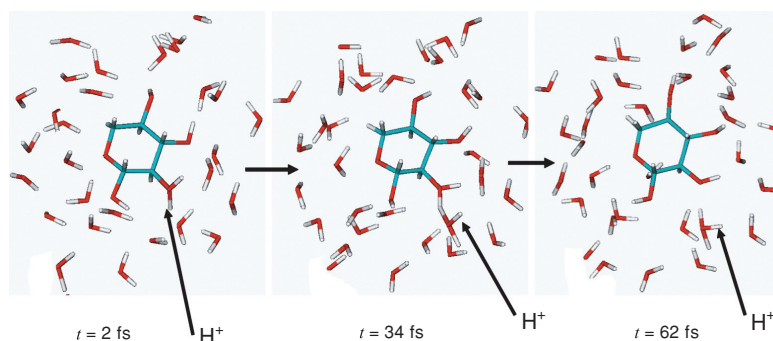


Plate 26 Snapshots of the MD simulations showing the rapid proton transfer from a xylose molecule to water. (See Figure 9.4.)

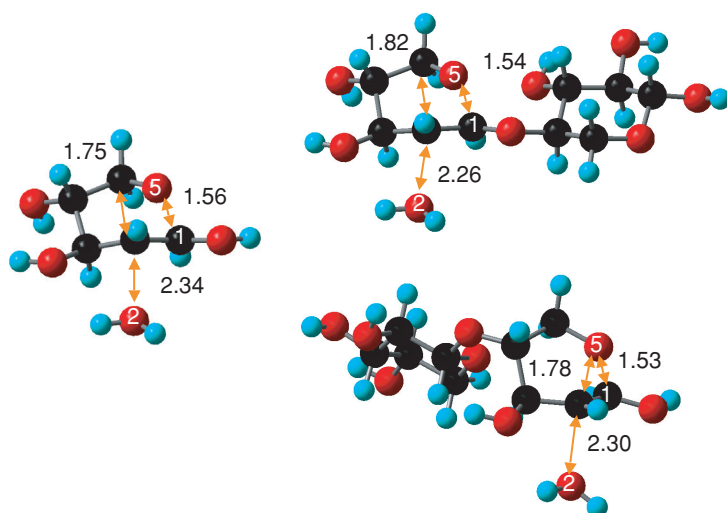


Plate 27 Transition states for dehydration reaction of xylose (left) and xylobiose (right-top and side view) resulting from protonation at O2. (See Figure 9.5.)

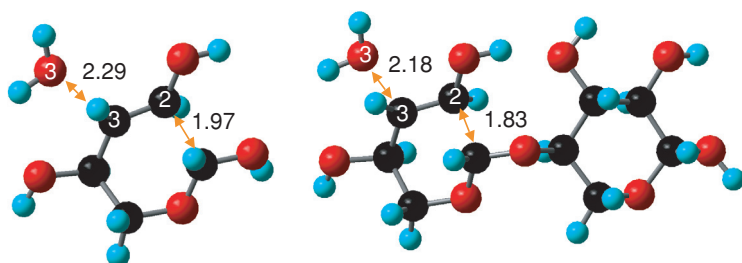


Plate 28 Transition states for dehydration reaction of xylose (left) and xylobiose (left) after protonation at O3. (See Figure 9.6.)

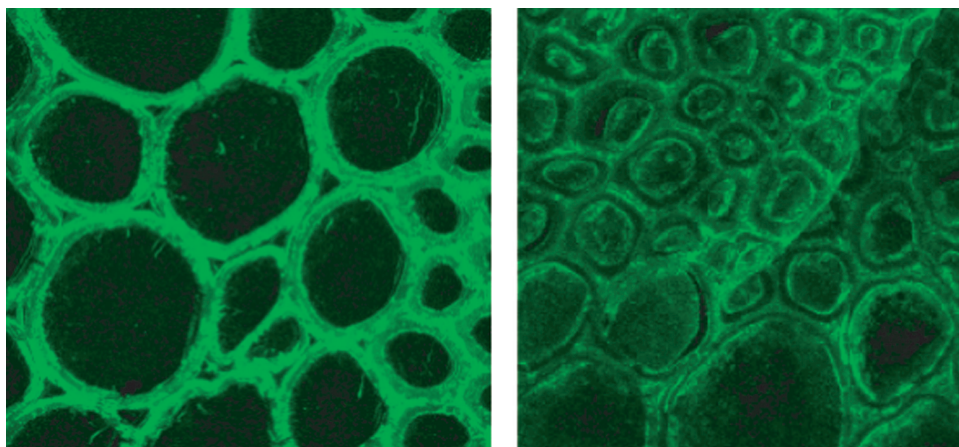


Plate 29 CLSM micrographs of unpretreated (left) and dilute acid pretreated (160°C, 10 min, 87% xylan removal) corn stover rind (right) sections labeled with LM11 α -xylan antibody that is bound to a secondary antibody conjugated to a fluorescent protein (Alexa488) showing decrease in fluorescent signal due to xylan hydrolysis (imaging by Stephanie Porter). (See Figure 14.2.)

Chapter 1

Our Challenge Is to Acquire Deeper Understanding of Biomass Recalcitrance and Conversion

Michael E. Himmel and Stephen K. Picataggio

1.1 The modern lignocellulose biorefinery

Alternative and renewable fuels derived from lignocellulosic biomass offer the potential to reduce our dependence on imported oil, support national economic growth, and mitigate global climate change (1, 2). However, breakthrough technologies are still needed to overcome barriers to developing cost-effective processes for converting biomass to fuels and chemicals. These needed breakthroughs include improved pretreatment processes that boost the yield of fermentable sugars while minimizing the formation and release of toxic byproducts; low-cost cellulases that hydrolyze crystalline cellulose; and microbial biocatalysts that enable rapid and efficient fermentation of the mixed sugars in cellulosic hydrolysates (3).

We also understand that feedstock costs will be a major component of the commodity end-product cost of biomass-derived liquid fuel products, such as ethanol and butanol. Therefore, perhaps the highest near-term priority is boosting the yield of lignocellulose-derived sugars. Yield issues touch many other critical biorefinery operations, including particle size reduction, pretreatment and detoxification, solids/liquid separation, enzyme hydrolysis, and fermentation of sugars to products.

It is now apparent that new process scenarios are also important for ensuring the success of future energy biorefineries. For example, the consolidation of existing process schemes may deliver significant economic and technical advantages. The well-known direct microbial conversion process proposed that a single microorganism could produce the cellulase enzymes and ferment sugars released from biomass to ethanol in high concentrations. Such a strain does not exist today, but could be constructed with suitable acquisition of a new, deeper understanding of various critical metabolic and enzymatic processes occurring in selected bacteria and yeast. Novel microbes may also allow a staged process to optimize these steps separately.

1.2 Biomass recalcitrance to deconstruction

We define the collective resistance that plants and plant materials pose to deconstruction from microbes and enzymes as “biomass recalcitrance.” This trait developed in terrestrial

plants during evolutionary maturation, in part, as a consequence of their moving from the protection of the aquatic environment.

Although little is known about the definitive steps involved or the intermediate forms explored, modern plants possess many systems for protection. The first line of defense in most plants is the epidermis, or outer layer of the plant anatomy. In grasses, this layer usually contains dense collections of thick-walled cells, as well as specialized cells that secrete waxy or oily materials. In trees, the bark presents a considerable physical, as well as chemical, barrier to all but the most dedicated assault.

Plant defense systems extend to the structure and organization of vascular tissue and even of the cell wall. Buried in the cell wall are the elementary fibrils that harbor the cellulose core (4). Even cellulose poses a significant barrier to enzyme action, where the highly ordered and water-excluding nature of the crystallite is sufficient to significantly retard cellulase action. This point is made especially clear when considering that the processive cellulase, cellobiohydrolase II, has been estimated from kinetic data to break about 14 bonds per second (5). Cell wall microfibrils are surrounded by sheaves of hemicellulose that, in turn, is covalently linked to lignins. This matrix of heteropolymers in which cellulose is embedded is certainly the dominant reason why plant biomass has resisted low-cost chemical and enzymatic treatment schemes.

1.3 Plants evolved to resist microbial and enzymatic assault!

We know that some plants, especially non-flowering ones, evolved rapidly during the Mesozoic Era. Ginkgos, for example, first appeared 150 million years ago and became common in the Mesozoic Era. One species, *Ginkgo biloba*, has been described as a “living fossil.” Certain characteristics enabled early plants to invade and become established on land. Internal vessels called vascular tissue circulated nutrients and water to all parts of the plant. An outer layer of waxy cuticle developed to prevent dehydration, and stomata located on the undersurfaces of leaves regulate respiration. Roots provide anchorage, nutrient uptake, and general interaction with the chemical/microbial systems in the soil (e.g., the rhizosphere).

New work to redirect the evolutionarily imposed protection of plants’ cell wall polysaccharides is now underway. The objective of “bioenergy plant engineering” is to use genetic tools to modify cell wall characteristics, thus permitting more-efficient chemical and enzymatic hydrolysis processes, as well as enhanced agronomic productivity. This work will proceed phenomenologically at first – for example, mapping plant quantitative trait loci to beneficial conversion traits. However, this field will mature to a deeper understanding of the processes of cell wall synthesis and assembly, as well as enzymatic deconstruction. Eventually, these biological systems will be sufficiently understood to permit overall system engineering, optimizing both cell wall production *and* deconstruction in ways not achievable in nature.

1.4 Are biomass-degrading enzymes working maximally?

Biomass-degrading enzyme preparations must work to convert as much of the polysaccharides in the cell wall as possible to monomers. Currently, high loadings of cellulases

are needed to reach 95% conversion of cellulose in pretreated biomass to sugars in 3–5 days using simultaneous saccharification and fermentation. Cellulase preparations are expensive in the biorefinery context for two reasons historically: 1) the source of the enzymes, usually *Trichoderma reesei*, was costly to grow and induce and 2) the specific performance (or activity) are low compared to other polysaccharide degrading enzymes.

However, a significant breakthrough in reducing the cost to produce and use *T. reesei* cellulases in the biorefinery was achieved by the DOE Office of the Biomass Program funded subcontracts awarded to Genencor International and Novozymes Biotech (2000–2005). Over the period of performance of these subcontracts, this cost was reduced about 10-fold from the starting cost of about \$5 per gallon of ethanol produced. We note that only a small percentage of the final cost reduction came from actually improving enzyme structure/function. The question is often asked: How low-cost must cellulases be to enable a new biorefinery industry? One answer may be based in considering the current cost of starch-degrading enzymes, which are about \$0.01–\$0.05 per gallon of ethanol produced. New amylase and glucoamylase technology introduced in 2005 will further lower these costs. For cellulase costs to approach that of starch-degrading enzymes, we must focus on considering the resistance of cellulose in plant microfibrils to deconstruction.

A deep understanding of the structure/function principles governing cell wall polysaccharidase action is critical. The study of cellulase action is especially challenging, considering that these enzymes function to first decrystallize cellodextrins and then hydrolyze the extracted chains to cellobiose and glucose. This process is not currently understood at the kinetic or thermodynamic level. Fundamental questions exist on the limits of enzyme activity and the action of soluble enzymes or enzyme aggregates on insoluble polymeric substrates in aqueous environments. It is possible, for example, that enzymes acting on microcrystalline cellulose are already working at the maximal rate!

The areas of poor scientific understanding presented above have clearly deterred past research programs aiming to reduce cellulase cost by improving performance. To summarize, the task of improving the specific activity of cellulases is complicated by our poor understanding of 1) cellulase natural diversity, 2) cellulase active-site architecture, 3) cellulase processivity, 4) cellulose decrystallization, and 5) the cellulose structure in plants.

1.5 Chemical pretreatments are still required to reveal cell wall cellulose

Thermal chemical pretreatments are currently necessary to enable cellulase access through the hemicellulose sheath of the plant cell wall microfibrils, thus exposing the crystalline cellulose core. This pretreatment must be just severe enough to create this access, but not so severe as to divert sugars to non-fermentable or toxic compounds (6–8). Today, the depolymerization of arabinoxylans (hemicellulose in hard woods and grasses) in cell walls is accomplished with good results by a variety of hot acid, hot water, and alkaline treatments. Final conversion of liberated soluble oligosaccharides is often accomplished using hemicellulases. Soft woods contain hemicellulose composed primarily of galactoglucomannan, which liberates galactose, glucose, and mannose upon depolymerization. These sugars are all fermentable by natural yeasts.

We recognize that the capital cost of the pretreatment unit operations is a critical factor for enabling the future biorefinery. High pretreatment capital cost is primarily due to the materials of construction required by conditions of high severity. In this context, severity is based on the pretreatment acidity, temperature, and time at temperature. New combinations of biological preconditioning (before thermal chemical pretreatment) and better thermal chemical pretreatments prior to enzymatic conversion have promise for overcoming this barrier.

The reactions of plant cell wall chemical constituents and ultrastructure to pretreatments must also be understood at a more detailed level. For example, basic research is required to understand the relationships between feedstock plant structure and composition. Simply, we need to develop better chemical and enzymatic treatments. Solving the yield challenge requires the integration of the complexities of plant structure, chemical pretreatment, and enzyme action. This integrated approach is a new and critical research paradigm.

1.6 Fermenting cell wall sugars: the stage is set for systems/synthetic biology

It is absolutely critical that the entire suite of sugars produced from all types of biomass be effectively converted to ethanol (or other products) by the fermentative microorganism – the ethanologen. A particular concern is the conversion of five-carbon sugars, primarily xylose and arabinose from grasses and hard woods. Desired characteristics for the ideal ethanologen include the following: it ferments all biomass sugars equally well (glucose, xylose, arabinose, galactose, mannose, and even sucrose); it resists toxic compounds produced during pretreatment (furfural, hydroxymethyl furfural, acetic acid, and soluble phenolics); it ferments high concentrations of sugars likely to be produced from high-solids pretreatments; and it produces fermentation beers with byproducts credits intact (9).

Realizing the potential of cellulosic biofuels may be facilitated by applying a new generation of genomic research tools. Metabolic engineering is now used routinely to develop microbial biocatalysts. Key approaches are the targeted manipulation of their metabolic pathways, or the introduction of new ones, with the goal of improving cellular properties or directing the synthesis of metabolic products with commercial value.

There are many examples now where metabolic engineering has improved the conversion yield, productivity, product concentration, and economic feasibility of an industrial bioprocess (10). One particularly relevant example is the success in extending the substrate utilization range of yeast and bacteria to include the pentose sugars derived from the hemicellulose fraction of biomass for conversion to fuel ethanol. Other examples include the introduction of genes that permit microorganisms to metabolize cellulose, starch, xylan, lactose, cellobiose, and sucrose. Other work has improved microbial growth rates and yields, nutrient uptake, and strain stability, or has reduced the overflow metabolism that causes the accumulation of inhibitory organic acid byproducts.

These efforts have provided a greater understanding of microbial physiology and the complexity of the interactions between metabolic pathways and their regulatory networks. A key discovery to emerge from these studies is that flux control is often distributed over several reactions in a pathway, rather than at a single “rate-limiting” step. Consequently, simultaneous and coordinated overexpression of all the genes encoding a metabolic pathway may

be necessary to increase metabolic flux without the detrimental accumulation of metabolic intermediates. Sophisticated *in-silico* models of complex metabolic networks are now used to define the minimal set of genes needed to optimize growth or product formation under particular conditions (11).

The “genomics revolution” has opened a whole new dimension to metabolic engineering. More than 800 microbial genomes have been sequenced thus far, representing enormous metabolic potential as a source of novel genes for strain development. Not too surprisingly, many enzymes catalyzing the same reaction in different microorganisms show widely varying kinetic properties. Furthermore, *in vitro* enzyme kinetics may not predict the *in vivo* activities of a complex pathway, making rational selection of best-pathway genes difficult. Combinatorial assembly of divergent homologs, coupled with strain selection and evolutionary adaptation, can overcome many of the limitations with rational gene selection.

The emerging field of synthetic biology now makes it possible to synthesize and assemble DNA fragments into modular cassettes that encode an entire metabolic pathway, synthetic chromosomes, and even whole genomes (12). The transplantation of a whole genome from one species of bacteria into another has recently been demonstrated and represents a major step toward developing customized microbial biofactories (13).

Microbial strain development historically relied almost exclusively on mutagenesis and selection to identify strains with superior traits, and the success of this approach is still evident today in the commercial production of amino acids, antibiotics, solvents, and vitamins. However, a systematic integration of the data generated by genomics, gene expression profiles, proteomics, and metabolomics offers the promise that we may develop a cohesive understanding of cellular metabolism sufficient to guide rational strain design. The new methods of synthetic biology now provide us with the means to introduce vast genetic diversity into a microbial host. And when combined with selection, high-throughput screening, and evolutionary adaptation, synthetic biology will allow us to identify those combinations of genes that optimize bioprocesses.

References

1. Perlack, R.D., Wright, L.L., Turhollow, A.F., Graham, R.L., Stokes, B.J. & Erblich, D.C. (2005) *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply*. DOE/GO-102005-2135. Oak Ridge National Laboratory, Oak Ridge, TN.
2. Foust, T.D., Wooley, R., Sheehan, J., Wallace, R., Ibsen, K., Dayton, D., Himmel, M., Ashworth, J., McCormick, R., Melendez, M., Hess, J.R., Kenney, K., Wright, C., Radtke, C., Perlack, R., Mielenz, J., Wang, M., Synder, S. & Werpy, T. (2007) *A National Laboratory Market and Technology Assessment of the 30x30 Scenario*. NREL/TP-510-40942. National Renewable Energy Laboratory, Golden, CO.
3. Himmel, M.E., Ding, S.Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W. & Foust, T.D. (2007) Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*, **315**, 804–807.
4. Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredes, A., Persson, S., Raab, T., Vorwerk, S. & Youngs, H. (2004) Toward a systems approach to understanding plant-cell walls. *Science*, **306**, 2206–2211.
5. Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T.T., Piens, K., Claeysens, M., Weber, M., Vasella, A., Becker, D., Sinnott, M.L., Zou, Z.-V., Kleywegt, G.J., Szardenings,

- M., Ståhlberg, J. & Jones, T.A. (2002) The active site of cellobiohydrolase Cel6A from *Trichoderma reesei*: The roles of aspartic acids D221 and D175. *Journal of the American Chemical Society*, **124**, 10015–10024.
6. Ding, S.Y. & Himmel, M.E. (2006) The maize primary cell wall microfibril: A new model derived from direct visualization. *Journal of Agricultural and Food Chemistry*, **54**, 597–606.
 7. McMillan, J.D. (1994) Pretreatment of lignocellulosic biomass. In: *Enzymatic Conversion of Biomass for Fuels Production* (eds. M.E. Himmel, J.O. Baker & R.P. Overend). American Chemical Society, Washington, DC.
 8. Wyman, C.E., Dale, B.E., Elander, R.T., Holtzapple, M., Ladisch, M.R. & Lee, Y.Y. (2005) Coordinated development of leading biomass pretreatment technologies. *Bioresource Technology*, **96**, 1959.
 9. Picataggio, S., Zhang, M. & Finkelstein, M. (1994) Development of genetically engineered microorganisms for ethanol production. In: *Enzymatic Conversion of Biomass for Fuels Production* (eds. M.E. Himmel, J.O. Baker & R.P. Overend). American Chemical Society, Washington, DC.
 10. Stephanopoulos, G., Aristidou, A. & Nielsen, J. (eds.) (1998) *Metabolic Engineering: Principles and Methodologies*, 725 pp. Academic Press, San Diego, CA.
 11. Burgard, A.P., Vaidyaraman, S. & Maranas, C.D. (2001) Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnology Progress*, **17**(5), 791–797.
 12. Pleiss, J. (2006) The promise of synthetic biology. *Applied Microbiology and Biotechnology*, **73**(4), 735–739.
 13. Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison, C.A., III, Smith, H.O. & Venter, J.C. (2007) Genome transplantation in bacteria: Changing one species to another. *Science*, **317**(5838), 632–638.

Chapter 2

The Biorefinery

*Thomas D. Foust, Kelly N. Ibsen, David C. Dayton,
J. Richard Hess, and Kevin E. Kenney*

2.1 Introduction

A significant challenge for the future will be meeting the world's mobility and chemical needs as populations and mobility needs grow. Currently, crude oil supplies almost all of the world's transportations' fuel need and a significant portion of the material and chemical needs (1). Overdependence on crude oil is leading to concerns about national energy security and short- and long-term price stability for both transportation fuels and commodity chemicals made from crude oil. Additionally, the everincreasing worldwide emissions of CO₂ from the transportation sector and the effect this is having on global climate change further increase concerns of over-reliance on crude oil.

Biomass, as the only source of renewable carbon, shows great promise for the large-scale economical production of renewable transportation fuels and chemicals. Biomass is an extremely abundant resource that can be produced in agriculture, forestry, and microbial systems. Biomass can also be captured from waste sources such as urban wood residues. Worldwide production of terrestrial biomass has been estimated to be on the order of 200×10^{12} kg (220 billion tonnes) annually. To put it in perspective, the total energy content from this amount of biomass (via heat of combustion analysis) is approximately five times the energy content of the total worldwide crude oil consumption (2).

A recent study (3) that looked specifically at US production capability showed the potential for the sustainable production of 1.3 billion dry tons per year of biomass from forest and agricultural lands without negatively impacting food, feed, and fiber production while still meeting export demands. Figure 2.1 shows the portion of this potential that could be produced from forest and agricultural lands by the middle of the next century with aggressive policies and economic incentives to maximize biomass production. The underlying assumptions in Figure 2.1 are that both agricultural resources and perennial energy crops are produced from agricultural land including some currently protected, or reserved, land with the forest resources being produced on forest land. Additionally, the agricultural resource potential includes grains that would be available for biofuels production.

Table 2.1 shows the transportation fuel production potential from the amount depicted in Figure 2.1. Ethanol was chosen as the representative biofuel because very accurate yield data exists for the various categories of feedstocks. To develop the yield numbers listed in Table 2.1, traditional corn dry-mill ethanol technology yields were used for the grain

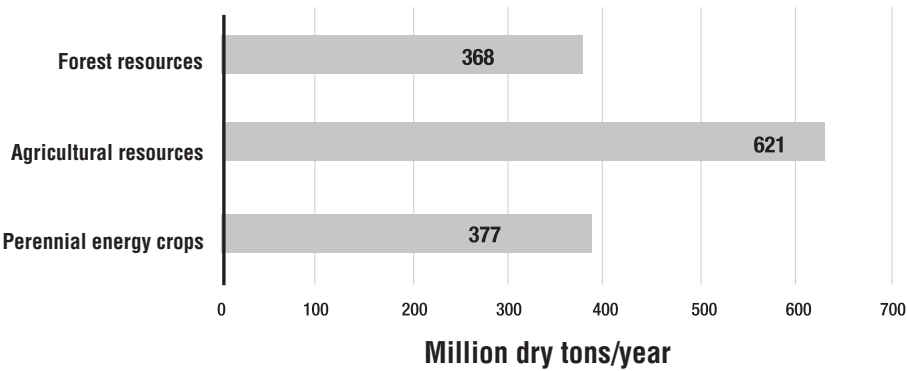


Figure 2.1 Agricultural and forest land resource potential.

resource potential, biochemical conversion process yields were used for the agronomic resource potential, and thermochemical process yields were used for the forestry resources. Biochemical- and thermochemical-to-ethanol yields are from Foust and coworkers (4).

In 2006, motor gasoline demand in the United States was approximately 143 billion gallons per year (5), so the ultimate potential for biofuels would be to supply approximately 60% of current gasoline demand on an energy basis. Although detailed data on worldwide biofuels potential does not exist, it is reasonable to assume the potential on a worldwide basis would be similar to the United States. However, only looking at biofuels as a replacement for existing transportation fuel usage is somewhat shortsighted and misleading. If the goal is to reduce dependence on crude oil, it is important to look at biofuels as a part of the overall solution to reducing worldwide dependence on imported oil. Scenarios have been developed in which sustainable development of biofuels, in conjunction with vehicle efficiency gains and smart growth, have been shown to be capable of virtually eliminating gasoline demand by the year 2050 (6). Given this significant potential, the challenge is to quickly deploy biofuels technology in an economically viable manner at a large scale.

Table 2.1 Total ethanol production potential

Resource	Tons available (million dry tons/yr)	Yield per ton of biomass (gal./ton biomass)	Total EtOH potential (billion gal./yr)
Grain ^a	89	104	9.2
Agriculture residues	532	90 gal./ton	47.8
Perennial energy crops	377	90 gal./ton	33.9
Forest resources	368	94 gal./ton ^b	34.6
Total – (volume basis)	1366		125.5
Total – (energy basis) – gallon gasoline equivalence (GGE)			84

^a Total grains available for biofuels production in high case from Perlack *et al.* (3) minus soybeans.

^b Total mixed alcohol production (80 gal./ton being EtOH).

The biorefinery concept is commonly presented as the method for large-scale deployment of biofuels. The term *biorefinery* was established in the 1990s (7) and has been refined many times over the years. The National Renewable Energy Laboratory's (NREL) web site defines biorefineries (8) as follows:

A biorefinery is a facility that integrates biomass conversion processes and equipment to produce fuels, power, and chemicals from biomass. The biorefinery concept is analogous to today's petroleum refineries, which produce multiple fuels and products from petroleum. Industrial biorefineries have been identified as the most promising route to the creation of a new domestic bio-based industry. By producing multiple products, a biorefinery can take advantage of the differences in biomass components and intermediates and maximize the value derived from the biomass feedstock while also being able to adapt to changing market conditions. The high-value products enhance profitability while the high-volume fuel helps meet national energy needs.

Sometimes in the literature the term "integrated biorefinery" is used (9) and is generally used in the context that a number of unit operations or technologies are used in an integrated manner to convert biomass to fuels and chemicals, so essentially the terms "biorefinery" and "integrated biorefinery" are interchangeable in the literature.

A good explanation of the stages or phases of biorefineries is provided by Kamm and Kamm (10) and Van Dyne and coworkers (11). They explain the progression of biorefineries in three phases as the technology develops to move from the simple, easily processed feedstocks at lower volumes to the more difficult to process lignocellulosic biomass feedstocks at higher volumes. An example of a Phase I biorefinery would be an existing corn dry-mill ethanol plant. It uses corn grain as the feedstock and a fairly straightforward set of conversion technologies that limit capital costs but put fairly confined constraints on production flexibilities and co-product production capabilities. In fact, most corn dry-mill ethanol plants are limited to an animal feed co-product, distiller's dry grain (DDG), as their only co-product in addition to ethanol. A dry mill has very limited capabilities to change its product mix to adapt to changing market conditions. However, a major advantage of corn dry-mill technology is its low capital cost. A cane sugar-based ethanol plant producing ethanol and food sugar, as currently exists in Brazil, is another example of a Phase I biorefinery.

An example of a Phase II biorefinery is an existing corn wet-mill plant. Although a wet-mill plant processes corn grain, it has more operational flexibility compared to a dry mill to produce a multitude of products such as ethanol, starch, high fructose corn syrup, corn oil, and corn gluten meal. The product mix in a corn wet mill can be varied to provide the highest economic return based on current market conditions. However, this co-product and product flexibility commands higher capital costs so wet mills tend to be twice as large, on average, compared to current dry mills to take advantage of economies of scale. Yet, even with the larger sizes, wet mills tend to have about a 10% higher capital cost per bushel of corn processed than dry mills (12). The Phase I and II biorefinery discussion does not need to be limited to corn wet and dry mills, an existing pulp and paper mill could be considered another example on a Phase I biorefinery that produces primarily a single project. A Phase III biorefinery would be essentially an integrated biorefinery capable of producing fuel(s) and other products from various feedstocks which would include lignocellulosic feedstocks.

Ethanol production has seen tremendous growth since 2000 primarily from new biorefinery construction and, to a lesser extent, from expanding the capacity of the existing

biorefinery industry (13). Essentially, all of the new ethanol production capacity being added is Phase I biorefineries, corn dry mills, suggesting investors favor lower capital costs over product flexibility. As an important historical reference, this indicates that lower capital costs should be favored over product flexibility for developing Phase III biorefineries.

Although corn grain and other starch- or sugar-based feedstocks processed in Phase I or II biorefineries can start the transition from sole dependence on petroleum for transportation fuels, as shown in Table 2.1, grain- and sugar-based biofuels have limited potential. Some studies (14) have listed the ultimate potential of biofuels considerably above the 9.2 billion gallons shown in Table 2.1, but the general consensus is that for biofuels to have significant production potential at the current scales of petroleum-based transportation fuels, the technology must be advanced to Phase III biorefineries that economically convert abundant lignocellulosic feedstocks into fuels and chemicals. The next sections describe the challenges and technology needed to achieve economic viability of Phase III biorefineries.

2.2 Phase III – lignocellulosic biorefineries

Given the variability of both the physical as well as chemical characteristics of the different feedstocks shown in Figure 2.2, robust conversion technologies must be developed that can economically accommodate this resource diversity. Additionally, feedstocks tend to be geographically diverse, i.e., softwoods in the Southeastern United States, and corn stover in the Midwest. Accommodating this diversity implies that the conversion technologies of future integrated biorefineries will be a function of the locally available feedstock resources. For this reason, cellulosic ethanol production technology must be sufficiently robust to optimize the conversion of multiple biomass resources to fuel.

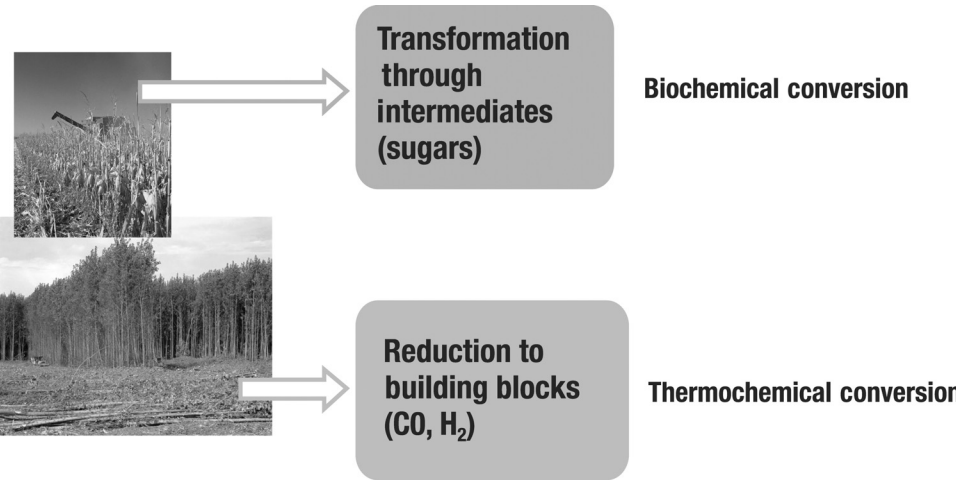


Figure 2.2 Primary conversion routes for cellulosic biomass.

Although there are a multitude of conversion technology routes under development for converting biomass to fuels and chemicals (15, 16), the predominant differentiation is the primary catalysis system (Figure 2.2).

Biochemical conversion uses biocatalysts (such as enzymes and microbial cells), heat, and chemicals to convert biomass first to an intermediate sugar stream and then to ethanol or other fermentation-produced fuel and co-products such as heat, power, and chemicals (17).

On the other hand, thermochemical conversion technologies use heat and/or physical catalysts to convert biomass to an intermediate product, and then a chemical transformation step to convert that intermediate product into fuels and chemicals. Thermochemical conversion technologies tend to be grouped into two distinct categories for fuel production (18): gasification and pyrolysis. Gasification conversion reduces biomass to a fundamental chemical building block, syngas (carbon monoxide and hydrogen), that can be reconstructed into ethanol and other fuel products through catalytic fuel synthesis processes (19). There are several gasification technologies capable of converting biomass to syngas via a network of chemical reactions, which can include partial oxidation, pyrolysis, and steam reforming, among others. There are several biomass gasification technologies under development that offer various pros and cons. A good recent review of the different types of biomass gasification technologies and their current state of development is provided by Spath and Dayton (19).

Pyrolysis is another thermochemical conversion technology but, unlike gasification that converts biomass to a syngas, pyrolysis converts the biomass to a liquid intermediate. Fast pyrolysis produces a pyrolysis oil or “bio-oil” in a short residence time process (0.1–2 s) in the absence of air at intermediate reaction temperatures typically in the range of 400–650°C. A good recent review on fast pyrolysis technologies and their current status is provided by Bridgwater and Peacocke (20, 21). In its produced form, the bio-oil is unacceptable for use directly as a transportation fuel because of its instability, high viscosity, and highly corrosive nature. Unfortunately, the highly heterogeneous nature of bio-oil makes the economical conversion to a fuel a very difficult challenge. However, many groups are currently researching the process of converting bio-oil to transportation fuels inside a petroleum refinery (22), and preliminary indications are that this work looks promising for the economical conversion of bio-oils to transportation fuel.

In addition to the base biological and thermochemical conversion routes of biomass to fuels listed above, there are a number of hybrid processes that take advantage of the synergies of both biochemical and thermochemical technologies for innovative and economically promising options for biorefineries. Some examples of such innovative approaches are syngas fermentation (23) and aqueous phase processes for converting sugars, sugar alcohols, and polyols into alkanes ranging from C₁ to C₁₅ (24–26). In fact, it can be argued that even the biological approach described in the next section is a synergistic biochemical and thermochemical approach in the fact that a thermochemical pretreatment process is the first step required in the process. Huber and coworkers (16) provide a good recent comprehensive review of biorefinery conversion technology options and their current status.

In the context of individual biorefineries there really is no clear single technology choice. Biorefinery developers will need to best match the conversion technology with the characteristics of the locally available feedstock(s) as well as match co-product options with local chemical markets to develop the best match for their particular set of conditions. Therefore, to realize the ultimate potential of biofuels for supplying the largest possible percentage of transportation fuels, the suite of conversion technologies must be capable of accommodating

the diversity of feedstocks. Hence, it is necessary to look closely at the feedstock conversion technology interface.

Thermochemical conversion technologies, particularly gasification approaches that reduce the biomass to a syngas, are robust to feedstock physical and chemical diversity (27). Forest residues – from small-wood forest thinnings or residues such as “hog fuel” from the forest products industry – are considered primarily for thermochemical conversion because of their compositional variability and lack of control over this diversity. Additionally, these types of feedstocks tend to have higher lignin content than herbaceous feedstocks, making them more suitable for thermochemical gasification conversion. Agricultural residues, in contrast, can be considered better suited for biochemical conversion technologies. These resources are expected to have a more uniform chemical composition because they are derived from cultivated crops that can be genetically engineered or selected for properties more amenable to biochemical conversion technologies (such as low recalcitrance or high cellulose or xylan content); this also holds for energy crops. The more uniform chemical composition of the feedstocks is in the macro sense, recognizing that in the micro sense there can be considerable variability. Biomass grown specifically for transportation fuel production can be engineered or selected to have the most desirable chemical and physical properties for a conversion technology. In addition, increasing the biomass resource base that can be biochemically converted to fuels provides an additional resource: lignin-rich fermentation residues that can be used for combined heat and power production or converted to biofuel in advanced, integrated biochemical–thermochemical biorefineries.

Although the long-term attractiveness of both biochemical and thermochemical lignocellulosic biomass conversion technologies looks very good, they are not yet economically competitive with either petroleum-derived gasoline or starch-based ethanol. To achieve economic viability of Phase III biorefineries, parallel efforts need to be undertaken to reduce both the feedstock cost component as well as the conversion cost component. The next three sections describe the R&D and technical challenges of achieving near-term economical competitiveness of Phase III biorefineries. The final section discusses long-term technology and R&D needs to realize the ultimate potential of biorefineries in supplying a significant portion of transportation fuel needs.

2.2.1 Feedstocks

The emerging biorefining industry is dependent on a large and sustainable supply of biomass resources provided at an effective cost and quality. In general, the feedstock cost can be subdivided into two components: the grower payment (or “stumpage fee” for forest resources) to cover the value of the biomass and the feedstock supply system costs. Feedstock supply system costs include harvesting, collecting, storing, handling, transporting, and any preprocessing required. Ultimately, market dynamics will control the grower payment component of the costs, and feedstock supply logistics will dictate supply system costs. Foust and coworkers (4) set a target of \$35/dry ton (2002 dollars) for the initial deployment of economically viable Phase III biorefineries in the United States, of which \$10 was for the grower payment. This is much too low to cover production costs for perennial energy crops. However, considering the 1.3-billion-ton potential, it is estimated that as much as 130 million tons could be accessed for a grower payment of less than or equal to \$10/dry ton, primarily from existing agricultural and forestry residues.

As the industry expands from grain ethanol to cellulosic ethanol, it is expected that agricultural crop and forest logging residues will be the first resources developed for biorefining. Energy crops will be integrated into the agricultural cropping system as the biorefining industry matures and creates a demand for them. An increase in energy crop production will likely occur as land managers (e.g., farmers and plantation foresters) use the additional crop options provided by the biomass energy market to maximize the productive capacity and economic returns of the land they manage.

The expanding use of lignocellulosic biomass resources will also create a demand for them, resulting in biorefineries paying more to access larger tonnages of the more expensive feedstocks (i.e., resources that require more than a \$10/dry ton grower payment). However, feedstock demand will always be limited by the price the biorefining industry can pay while remaining competitive in the ethanol fuel market. Initially, government policies and programs may be the means to access higher-value feedstocks. Up to and beyond the 2030 time frame, technology advancements will reduce feedstock supply system costs, which will then provide increased purchasing power for biorefineries to access higher-value biomass feedstocks. This strategy of improving supply and conversion technologies to purchase higher-value feedstocks is well established in other processing and refining industries (28). This combination of policy and technology advancement will help develop a biomass resource large enough to support the long-term goals of producing economical biofuels on a large scale.

2.2.1.1 Feedstock R&D pathway

The biorefinery feedstock supply system encompasses all the unit operations necessary to move biomass feedstocks from the land to the biorefinery (29). An overview of the feedstock supply system is depicted in Figure 2.3. Biomass production is the beginning of the feedstock

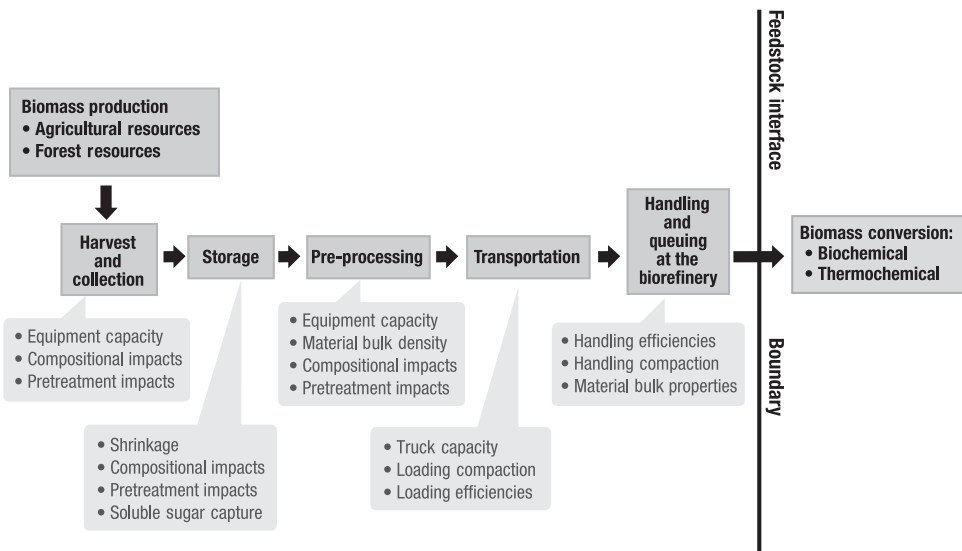


Figure 2.3 Feedstock supply system schematic including key barriers to economic viability.

supply chain. It involves producing biomass feedstocks to the point of harvest. Production addresses important factors such as selection of feedstock type, land use issues, policy issues, and agronomic practices that drive biomass yield rates and directly affect harvest and collection operations.

Harvesting and collection encompasses all operations associated with getting the biomass from its source to the storage or queuing location. In addition to obvious operations such as cutting (or combining, swathing, or logging) and hauling, this often includes some form of densification such as baling, bundling, or chipping to facilitate handling and storage. Storage and queuing are essential operations in the feedstock supply system. They are used to deal with seasonal harvest times, variable yields, and delivery schedules. The objective of a storage system is to provide the lowest-cost method (including cost incurred from losses) of holding the biomass material in a stable, unaltered form (i.e., neither quality improvements nor reductions) until it is called for by the biorefinery.

Prior to conversion, the feedstock must be preprocessed to physically transform it into the format required by the biorefinery. Preprocessing can be as simple as grinding and preparing the biomass for increased bulk density or improved conversion efficiency, or it can be as complex as improving feedstock quality through fractionation, tissue separation, and blending.

Transportation generally consists of moving the biomass from the storage location to the biorefinery via truck, rail, barge, or pipeline. The system used directly affects how the feedstock is handled and fed into the conversion process. Transporting and handling methods are driven by the format and bulk density of the material; this makes them highly dependent on each other and on all the other operations in the feedstock supply chain.

2.2.1.2 R&D needs to achieve economic viability

Significant advances have been made to transform the feedstock supply process from traditional technologies historically used in smaller distributed livestock, forage, and wood product industries to an assembly system specifically designed for the biorefinery industry.

Feedstock infrastructure development is difficult because equipment, methods, and logistics vary not only among resources (e.g., agricultural residues versus forest residues), but also among geographic regions (e.g., dry agricultural residues in dryer regions versus wet agricultural residues in wetter regions). Consequently, the feedstock supply infrastructure must be developed for each class of biomass resource. Three general classes that cover the range of feedstocks are as follows:

- Dry herbaceous (examples: stover, straw, and switchgrass that are harvested at <15% moisture dry basis by weight). Dry herbaceous feedstocks present the fewest logistical challenges for use as biorefinery feedstocks. Although limited in overall volume, they provide a good opportunity for near-term utilization for establishing Phase II biorefineries.
- Wet herbaceous (examples: stover and switchgrass that are harvested at >15% moisture dry basis by weight). The use of wet herbaceous feedstocks is limited by a host of infrastructure barriers. Because wet herbaceous feedstocks represent a significant portion of the overall feedstock resource, overcoming these barriers provides the greatest potential to achieve the projected tonnage targets.

- Woody (example: logging residues). Logging residues have been used for energy in Europe and the United States for nearly 30 years. As a result, the logging residue supply system is quite mature, and systems and methods are already developed to support this industry (30). Because near-term woody feedstock will consist largely of logging residues, the infrastructure for this feedstock can be readily adapted and validated against resource environment, resource policy, and other regional factors.

The R&D activity plan for developing and validating a feedstock supply infrastructure capable of producing large tonnages of biorefinery feedstocks at the lowest possible cost addresses one of three key factors – equipment capacity, equipment efficiency, or feedstock quality – affecting feedstock supply system costs. The specific research plan that focuses on the application of these cost factors to each of the supply system elements for achieving the feedstock R&D targets is described below.

2.2.1.2.1 PRODUCTION

Production is a critical component of the feedstock supply system, and it is a key component to ensuring an adequate and sustainable feedstock supply. Specific research needed to address production issues includes:

- Assessing the cost and availability of the feedstock resource on a local basis to define production costs (e.g., grower payments) and identify regional tonnages available within each feedstock type or classification at or under the feedstock threshold costs
- Identifying and validating sustainable agronomic and silviculture practices specific to feedstock types and regional variables to ensure sustainable production
- Investigating crop production improvements (e.g., increased yields, decreased yield variability, and consistent quality) through genetic modification
- Developing a perennial crop program that includes matching varieties to site conditions, establishing optimum agronomic and silviculture practices, and developing a seed production program.

2.2.1.2.2 HARVEST AND COLLECTION

Harvest and collection advances are required in three key areas: 1) selective harvest (including forest thinning operations), 2) single-pass or minimum-impact harvest, and 3) harvest and collection efficiencies. The primary drivers for improved harvest technologies are reduced costs and access to larger tonnages of biomass through increased producer participation. For example, improved harvest technologies that address soil quality concerns – such as carbon sequestration, nutrient/water retention, erosion, and compaction will become increasingly important for enticing grower participation and accessing biomass resources.

Performance metrics for new harvest and collection systems include: 1) efficiency, 2) equipment capacity (an element of efficiency that includes technologies that reduce capital and improve throughput of equipment), and 3) quality. Without these improvements, the accessible biomass tonnage remains restricted.

Needed research in this area includes:

- Developing innovative harvest and collection methods for all resource types to eliminate or reduce unit operation costs and agronomic silviculture operational impacts

- Understanding, quantifying, and validating harvesting-specific quality related to compositional effects, pretreatment effects, contaminant reductions, and bulk handling improvements
- Developing and testing innovative equipment specific to woody feedstocks for which existing equipment is too costly and inefficient.

2.2.1.2.3 PREPROCESSING

Recently, there have been significant R&D advances in dry herbaceous preprocessing that will enable the transition from the current state-of-technology bale-based system to a more cost-effective bulk feedstock system for biorefineries. However, additional advances are needed in three key areas: 1) preprocessing equipment capacity, 2) feedstock bulk density, and 3) feedstock quality. Equipment capacity and bulk density directly affect feedstock cost. Thus, they are important technical parameters to address, along with the interrelated effect on feedstock rheological properties. Furthermore, a key component of feedstock R&D is to extend preprocessing beyond size reduction to include value-added operations that improve feedstock quality for the biorefinery, such as fractionation and separation of higher-value components.

Specific research needed in this area includes:

- Developing preprocessing requirements for each feedstock type
- Understanding the relationship between biomass structure and composition for assessing quality-upgrade potential and developing equipment and methods to achieve these upgrades
- Understanding and controlling biomass tissue deconstruction in preprocessing and the relationships among grinder configuration, tissue fractions, tissue moisture, and grinder capacity to optimize grinder configuration for fractionation, capacity, and efficiency
- Increasing bulk densities by coupling the understanding of biomass deconstruction and rheological properties with innovative bulk compaction methods
- Understanding and controlling feedstock rheological properties resulting from preprocessing operations to provide a product that minimizes problems in transportation, handling, and queuing operations.

2.2.1.2.4 STORAGE AND QUEUING

Feedstock shrinkage (or dry matter loss) and quality reductions are major problems during feedstock storage. Shrinkage and quality reduction risks and mitigation strategies vary widely from region to region. Although rigorously developed targets for dry matter losses have yet to be developed, the general consensus is that they must be less than 5% for all feedstock types.

Specific research needed in this area includes:

- Assessing storage options and their effects on dry matter losses, compositional changes, and functional biomass changes specific to resource type and regional variables
- Establishing baselines of storage systems costs at scales from 0.8 million tons/year to 10 million tons/year to identify key cost and infrastructure issues and develop paths to minimize industrial-scale storage costs

- Understanding soluble sugar and carbohydrate loss and evaluating the feasibility of preventing or reclaiming those soluble sugars and carbohydrates from the feedstock during storage
- Developing cost-effective methods of large-scale bulk storage that reduce handling, eliminate bulk flow problems, and minimize adverse physical changes that may affect plant processing.

2.2.1.2.5 TRANSPORTATION AND HANDLING

Transportation is a significant cost, and it can be a barrier to using some feedstock resources. Transportation and handling operations can account for nearly 50% of the capital investment of a feedstock assembly system. Unlike the other unit operations in the feedstock supply system that can impart additional value to the feedstock, transportation costs simply move the feedstock to the biorefinery. Hence, reducing these costs to a minimum is vital to achieving low feedstock supply system costs.

Regardless of the transport method (e.g., truck, rail, or barge), bulk density is the key technical parameter that must be addressed to decrease transportation costs. As such, methods to increase bulk density are a focus of the transportation and handling R&D. Bulk handling is also affected by feedstock rheological properties and this, too, is an area of focus.

Specific research needed to reduce transportation and handling costs includes:

- Understanding physical and rheological feedstock properties (including bulk density) as they relate to handling systems to optimize handling and transportation efficiencies
- Evaluating innovative transportation and handling methods.

2.2.2 Biochemical conversion

2.2.2.1 Introduction

Basically, biochemical conversion is the liberation and fermentation of sugars from biomass feedstocks. The challenge is to efficiently convert the carbohydrate portion of the biomass to sugars, or “saccharify” it, and ferment the impure sugars to ethanol with a robust microorganism. In this process, the lignin component of the biomass provides the heat and power needs of the process. This process shows great promise for producing ethanol cost effectively with high yields and minimal environmental impact.

There are two primary routes for saccharification: 1) acid hydrolysis, with either concentrated or multiple stages of dilute; and 2) pretreatment followed by enzymatic hydrolysis. In the 1980s, DOE evaluated the long-term potential of each process (31) and although at the time acid hydrolysis technology was further developed and appeared less expensive, comparing progress and future potential suggested that enzymes offered greater opportunity for ethanol cost reduction in the long run (32). Acid hydrolysis technologies are certainly feasible and in proper niche situations they are being pursued to commercialization.

Enzyme hydrolysis requires a pretreatment to generate an intermediate material that can be effectively digested by enzymes. Dilute acid pretreatment of corn stover followed by enzymatic hydrolysis can achieve more than 90% conversion of cellulose to glucose (33). Various pretreatment methods have been suggested; most use heat coupled with a chemical catalyst such as an acid, base, or other solvent. Recent advances (34) suggest that “accessory”

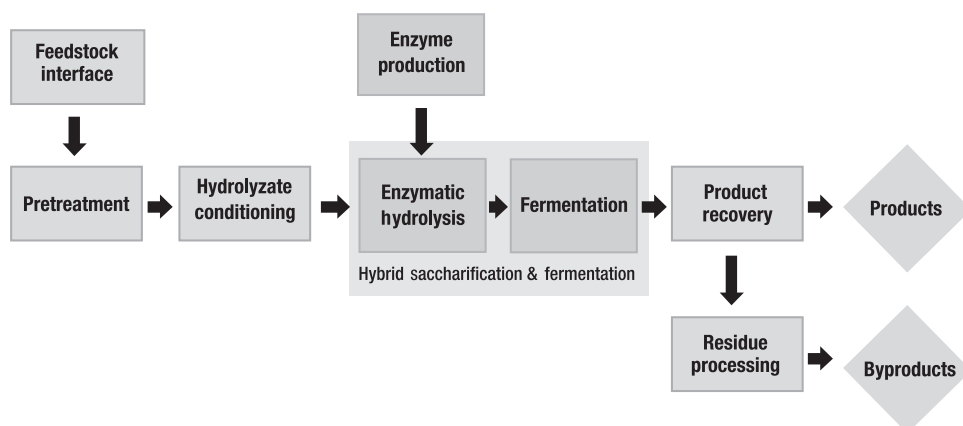


Figure 2.4 Process schematic for biochemical conversion.

enzyme systems such as hemicellulases could lead to low-severity and low-cost pretreatment processes in the future. Although currently it appears that dilute-acid-based approaches give the best overall performance over the range of feedstocks envisioned for biochemical conversion, other approaches such as alkaline approaches also show considerable promise; and more development is needed in the pretreatment area to meet cost performance goals. Wyman and coworkers (35) give a good recent review on the comparative performances of the leading pretreatment technologies under development.

A representative block flow diagram of a biochemical conversion route to convert lignocellulosic biomass to ethanol using dilute acid pretreatment followed by the enzymatic hydrolysis approach is shown in Figure 2.4. The process also includes ancillary supporting operations such as feedstock interface handling and storage, product recovery, wastewater treatment, residue processing (lignin combustion), and product storage not shown in Figure 2.4.

The feedstock is delivered to the feed-handling area for size reduction and storage. From there, the biomass is conveyed to pretreatment and conditioning. In this area, the biomass is treated with a dilute sulfuric acid catalyst (the current leading pretreatment technology) at a high temperature for a short time. This hydrolyzes the hemicellulose to a mixture of sugars (i.e., xylose, arabinose, galactose, mannose, and a small amount of glucose) and other compounds. In addition, the pretreatment step makes the remaining biomass more accessible for later enzyme saccharification. A conditioning process then removes byproducts from the pretreatment process that are toxic to the fermenting organism.

In hybrid saccharification and co-fermentation (HSF), the pretreated solids (now primarily cellulose) are saccharified with cellulase enzymes to form monomeric glucose. This requires a couple of days, after which the mixture of sugars and any unreacted cellulose is transferred to a fermenter. An inoculum of fermenting microorganism is added, and the sugars are fermented to ethanol. Meanwhile, the enzymes are used for further glucose production from any remaining biomass, which is now at conditions optimal to fermentation. After a few days of fermentation and continued saccharification, nearly all the sugars are converted to ethanol. The resulting beer (or low-concentration ethanol) is sent to product recovery.

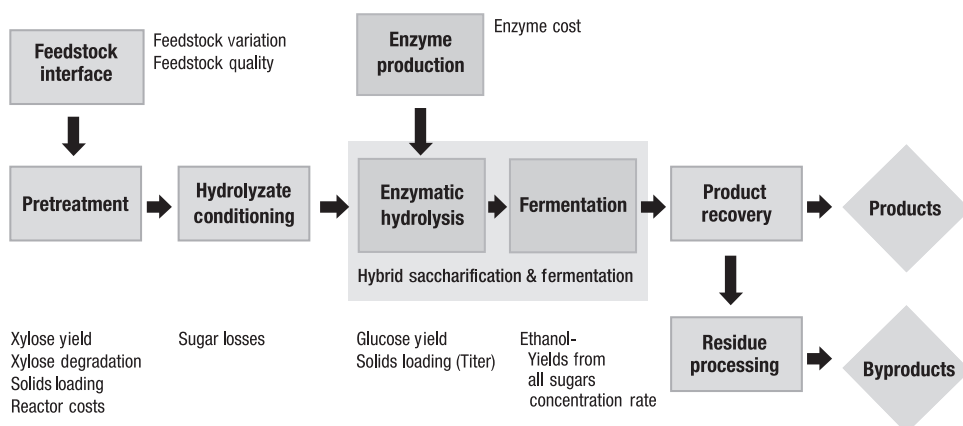


Figure 2.5 Process flow diagram highlighting major research barriers.

Product recovery involves distilling the beer to separate the ethanol from water and residual solids. An azeotrope of water and ethanol is converted to pure ethanol using vapor-phase molecular sieves. Solids from the distillation bottoms are separated and sent to the boiler (called residue processing). Distillation bottoms liquid is then concentrated by evaporation using waste heat. The evaporated condensate is returned to the process, and the concentrated syrup is sent to the burner.

Part of the evaporator condensate, along with other wastewater, is treated by anaerobic and aerobic digestion. The biogas (which is high in methane) from anaerobic digestion is sent to the burner for energy recovery. The treated water is suitable for recycling and is returned to the process.

The solid distillates – the concentrated syrup from the evaporator and biogas from anaerobic digestion – are burned in a fluidized bed combustor to produce steam for process heat. The majority of the steam demand is in the pretreatment reactor and distillation areas. Generally, the process co-generates enough electricity to use in the plant and to sell to the grid. A detailed description of the conversion process described above is provided by Aden and coworkers (17).

2.2.2.2 R&D needs to achieve economic viability

Accomplishing economic competitiveness with gasoline and starch-based ethanol requires additional technology advancement in key areas of pretreatment, enzymatic hydrolysis, and fermentation processes (4). Figure 2.5 shows the technical barriers that must be addressed for the individual unit operations to significantly reduce overall process costs. The following sections describe the research and development needed to overcome these barriers for the individual unit operations as well as integration challenges for the overall process.

2.2.2.2.1 FEEDSTOCK/PROCESS INTERFACE R&D NEEDS

As stated above, the biological conversion process is highly sensitive to the physical and chemical characteristics of the feedstock. Because it is unlikely that the process can be

supplied with a standardized homogeneous feedstock for continuous year-round operation, it will be necessary to adjust and modify the process to accommodate feedstock variations. To accomplish this, researchers will need to:

- Understand the range of feedstock types expected to be used in pioneer and subsequent plants
- Work with feedstock suppliers and researchers to improve the quality (physical and chemical characteristics) of the feedstocks
- Determine the impacts of the feedstocks on downstream unit operations
- Understand how to adjust the process to maintain optimal yields and productivities with varying feedstock quality or different feedstocks.

2.2.2.2.2 PRETREATMENT AND HYDROLYZATE CONDITIONING R&D NEEDS

A major goal of pretreatment and hydrolyzate conditioning is to maximize the xylan-to-xylose conversion while minimizing the xylan lost to degradation products. Additionally, to maximize plant throughput per installed capital, it is desirable to operate at the highest levels of solids loading as possible without compromising performance goals. Specific research needs to accomplish these objectives are as follows:

- Determine the location of the xylan in the plant cell wall and optimize pretreatments that selectively remove and hydrolyze it to xylose
- Reduce sugar degradation to minimal levels by understanding the kinetic mechanisms that lead to undesirable degradation products and then investigating methods to systematically block these pathways.

A second major goal is to reduce the capital cost of pretreatment by using ancillary enzymes. The research goal here is to determine if other enzymes, such as xylanases, can improve xylose yields, minimize the formation of degradation products while reducing the costs associated with the pretreatment process. Another challenge facing the biochemical conversion processes described by Aden and coworkers (18) is the need to condition the hydrolyzate to reduce the toxicity prior to fermentation. Aden *et al.* identifies overliming as the most promising current process for conditioning the hydrolyzate prior to fermentation. However, overliming is problematic because it adds an additional unit operation as well as an associated solid–liquid separation step that increases both the capital and operating costs as well as leading to significant sugar losses associated with the overliming process (4). There are some pretreatment approaches (36) being tested that show considerable promise for eliminating or greatly reducing the need for extensive conditioning that would be highly beneficial from a cost, yield, and overall process simplicity perspective. For any pretreatment approach under development, it is highly desirable to eliminate or greatly reduce the need for conditioning. Reducing the severity of pretreatment to reduce degradation product formation and increasing the robustness of fermentation organisms will contribute to eliminating or significantly reducing conditioning needs. Specific research needed to accomplish these objectives is as follows:

- Understand the role of hydrolyzate conditioning to eliminate sugar losses
- Understand and control the degradation kinetics to minimize or eliminate the formation of inhibitory compounds.

2.2.2.2.3 ENZYME PRODUCTION R&D NEEDS

Over the years, significant progress has been made in reducing cellulose enzyme costs, and very impressive cost reductions have been reported recently (37, 38). Although this progress is a major step toward the economic competitiveness of the overall biochemical process, further cost reductions in cellulose enzymes are still needed. Essentially there are two ways to reduce cellulose enzyme costs, which are typically expressed on a normalized cost of gallon-of-ethanol-produced basis: increase enzyme-specific activity or decrease enzyme production costs. Although both enzyme cost reduction approaches are useful and necessary to accomplish the cellulase enzyme cost reduction goal, increasing the enzyme-specific activity has the added benefit of reducing saccharification time, which also increases the effective utilization of capital. Specific research needed to accomplish these objectives is as follows:

- Understand cellulase interactions at the plant cell wall ultrastructural level to optimize hydrolysis processes, enzyme kinetics, and, ultimately, cellulase use and cost
- Determine how cellulase enzymes move along the cellulose chain and the roles of enzyme substructures
- Conduct targeted substitutions of enzyme components to increase specific activity guided by molecular modeling of cellulase/substrate interactions
- Identify enzyme production processes and logistics to minimize processing and transportation costs of enzyme products.

2.2.2.2.4 FERMENTATION R&D NEEDS

Developing a robust, commercially viable biocatalyst (or microorganism) capable of fermenting a large percentage of both the hemicellulose sugars as well as glucose to commercially viable ethanol titers at commercially viable fermentation times is a significant goal. Foust and coworkers (4) define near-term commercially viable targets as capable of fermenting 85% of hemicellulose sugars and 95% of glucose to a concentration of at least 6% ethanol in 3 days in combined hybrid saccharification and fermentation, while maintaining the solids concentration necessary for the ethanol concentration target above. Specific research needed to accomplish these objectives is as follows:

- Identify strain candidates that exhibit superior “wild-type” performance
- Use metabolomics, proteomics, and other tools to understand metabolic bottlenecks in the carbon assimilation pathways that limit pentose sugar uptake and the ability to withstand fermentation inhibitors such as organic acids, low pH, and increased temperature
- Extend “omics” studies to identify and understand secondary pathway limitations related to reaction cofactors and regulation of metabolism
- Increase pentose uptake rates by applying protein and metabolic engineering to increase sugar transporter efficiency, pentose specificity, and expression
- Improve strain robustness by manipulating cell membrane composition to reduce its permeability to organic acids and improve its temperature stability
- Use a combination of metabolic engineering, mutagenesis, and long-term culture adaptation strains on actual pretreatment hydrolyzate to achieve targeted fermentation performance

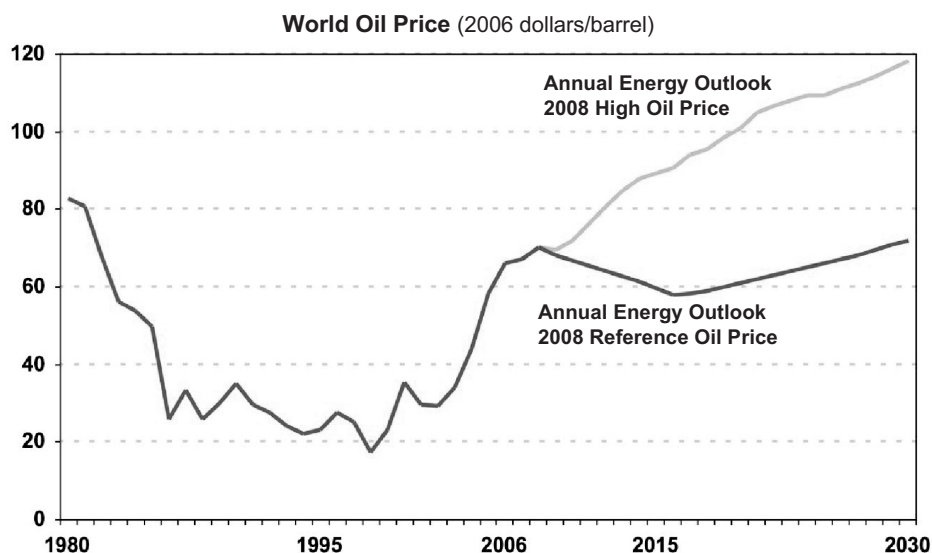


Figure 2.6 World oil price scenarios.

- Perform parametric analysis of such factors as lignin redeposition and the detrimental effects this can have on enzyme kinetics to minimize these effects
- Use information about the enzyme capabilities and fermenting strain's performance to develop and test strategies for efficiently integrating enzymatic hydrolysis with biomass sugar fermentation to maximize cellulose hydrolysis and sugar fermentation rates and yields
- Quantify the effects of enzyme loading, strain inoculation time, and inoculum charge on batch process performance
- Use reactor designs and operational schemes to maximize the solids loading and conversion of cellulose and other carbohydrates to ethanol.

2.2.2.2.5 INTEGRATION/PROCESS ENGINEERING R&D NEEDS

Finally, it is important to consider all the individual unit operations in the context of an integrated process. Although there is general agreement in the literature about necessary performance targets for individual unit operations as outlined above, overall integrated process targets are more difficult to define. Essentially, the overall process must be economically viable in the fuel marketplace, and the process must be demonstrated at some reasonable pilot scale of continuous reliable operation. Economic viability in the fuel marketplace is largely a function of gasoline prices, which are driven by crude oil price projections. Figure 2.6 shows the latest DOE Energy Information Agency 2008 Annual Energy Outlook projections for world crude oil prices out to 2050.

Foust and coworkers (4) picked \$1.07 in 2002 dollars as a production cost target, which includes an IRR (internal rate of return) of 10% as an economically viable target for cellulosic ethanol. The rationale provided for this target was based on historical fuel ethanol prices as shown in Figure 2.7. The \$1.07 in 2002 dollars (\$1.31 in 2007 dollars) per gallon value

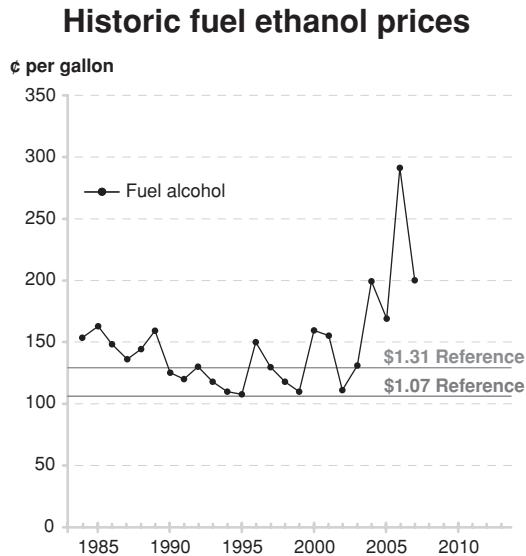


Figure 2.7 US list prices for ethanol.

represents the low side of the historical fuel ethanol prices, and, given historical price data, cellulosic ethanol would be commercially viable at this production cost.

In addition, the target price of \$1.31 per gallon of ethanol is also in line with current gasoline rack, or pre-tax, prices. To compare the target ethanol price with the price of gasoline on an “apples to apples” basis at the pump, the ethanol price must be adjusted as follows:

- 1 Adjust the ethanol price from dollars per gallon of ethanol to dollars per gallon of gasoline equivalent by correcting for the two-thirds lower energy content of ethanol compared to gasoline. This increases the \$1.31 per gallon ethanol to \$1.96 per gallon gasoline equivalent.
- 2 Adjust ethanol price from plant gate to retail price. The price of gasoline includes, on average, \$0.40 per gallon for taxes and \$0.23 for distribution. Assuming the same costs for ethanol gives it a retail price slightly higher than that of gasoline when oil is at \$65 per barrel, as shown in Figure 2.8. This price at the pump analysis does not assume any subsidy for ethanol.

2.3 Thermochemical biorefinery

2.3.1 Introduction

Thermochemical conversion technology options include gasification and pyrolysis. Although both processes show long-term promise, gasification approaches show considerable promise for near-term economically competitive liquid fuels production. As stated

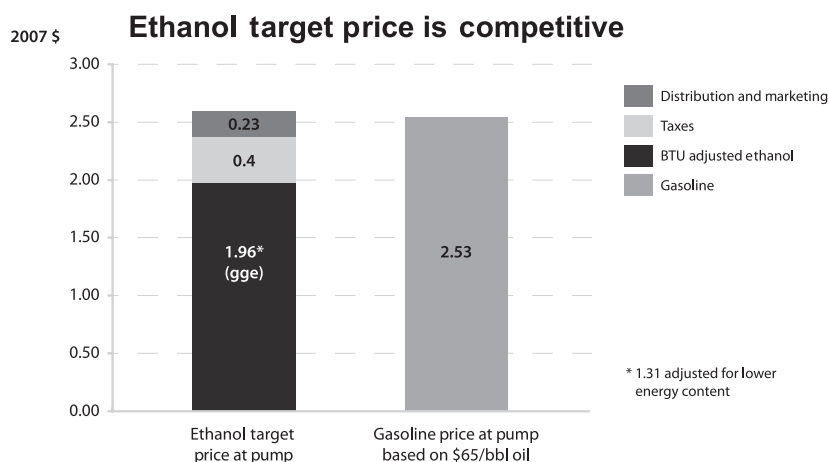


Figure 2.8 Ethanol and gasoline price comparison.

above, the thermochemical process to liquid transportation fuels adds technology robustness to a scenario for producing a significant portion of transportation fuels from biomass. It can convert low-carbohydrate, or “non-fermentable,” biomass materials such as forest and wood residues to fuels at lower technical challenge levels than the biochemical conversion process route. This section describes the R&D needed to achieve economical competitiveness for a stand-alone biomass gasification/mixed alcohol process.

Biomass gasification converts heterogeneous feedstock supplies into a consistent gaseous intermediate that can be converted to liquid fuels. The product gas called “synthesis gas” or “syngas” has a low-to-medium energy content (depending on the gasifying agent) and consists mainly of carbon monoxide, hydrogen, carbon dioxide, water, nitrogen, and hydrocarbons. Minor components, also referred to as contaminants, include tars, sulfur and nitrogen oxides, alkali metals, and particulates. These contaminants threaten the success of downstream syngas to liquid fuels conversion and must either be reformed or removed. Commercially available and near-commercial syngas conversion processes were evaluated on technological, environmental, and economic bases by Spath and Dayton (19). Their report provides the basis for identifying promising, cost-effective fuel synthesis technologies that maximize the impact of biomass gasification.

Figure 2.9 shows a representative block process flow diagram of a thermochemical process that produces ethanol from lignocellulosic biomass. The process also includes ancillary supporting operations such as feedstock interface handling and storage, product recovery, and product storage not shown in the figure. Phillips and coworkers (39) provide a detailed description of this process, which is being capable of producing economically viable ethanol from a plant processing 2000 tonnes/day of biomass. Although syngas-to-liquid processes are capable of producing a variety of transportation fuels, ethanol was selected here to provide synergy with the biochemical conversion route. A brief overview of the process developed by Phillips and coworkers is described below.

The feedstock interface addresses the main biomass properties that affect the long-term technical and economic success of a thermochemical conversion process: moisture content, fixed carbon and volatiles content, impurity concentrations, and ash content. High moisture

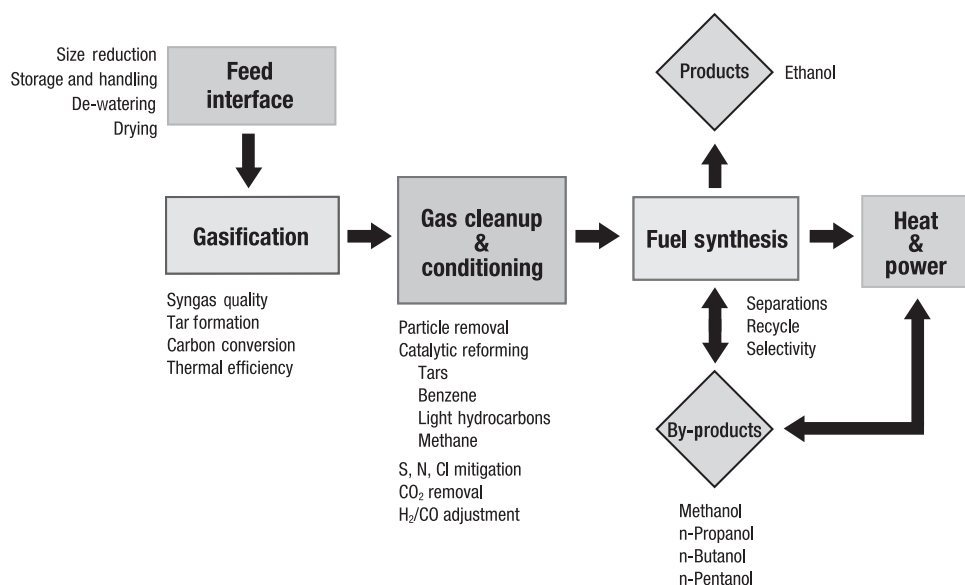


Figure 2.9 Process flow diagram with research barriers for economical thermochemical ethanol production.

and ash content reduce the usable fraction of delivered biomass. Therefore, maximum system efficiencies are possible with dry, low-ash biomass.

Gasification is a thermochemical process that involves the thermal decomposition of biomass at temperatures that maximize syngas yield. Tar and char produced during decomposition may also react with steam, CO₂, and hydrogen in the gasifier to produce additional gas. This is followed by partial oxidation of the fuel with a gasifying agent – usually air, oxygen, or steam – to yield raw syngas. The raw gas composition and quality are dependent on a range of factors, including feedstock composition, type of gasification reactor, gasification agents, stoichiometry, temperature, pressure, and the presence or lack of catalysts.

Gas cleanup removes contaminants from biomass gasification product gas. It generally involves an integrated, multi-step approach, which varies depending on the intended end use of the product gas. However, gas cleanup normally entails removing or reforming tars, acid gas removal, ammonia scrubbing, capturing alkali metal, and removing particulates. Gas conditioning is the final modification to gas composition to make it suitable for a fuel synthesis process. Typical gas conditioning steps include sulfur polishing (to reduce hydrogen sulfide to acceptable levels for fuel synthesis) and water–gas shift (to adjust the final hydrogen–carbon monoxide ratio for optimized fuel synthesis).

Comprehensive cleanup and conditioning of the raw biomass gasification product gas yields a “clean” syngas composed of carbon monoxide and hydrogen in a given ratio. This gas can be converted to a mixed-alcohol product. Separation of the ethanol and higher molecular weight alcohols from this product yields a methanol-rich stream that can be recycled with unconverted syngas to improve process yield. The higher-alcohol-rich stream yields byproduct chemical alcohols. The fuel synthesis step is exothermic, so heat recovery is essential to maximizing process efficiency.

2.3.2 R&D needs to achieve economic viability

As is the case with the biochemical conversion route discussed in the previous section, accomplishing economic competitiveness with gasoline and starch-based ethanol requires additional technology advancement in key areas of gas cleanup, conditioning, and fuels synthesis (4). The following sections describe the R&D needed to overcome these barriers for the individual unit operations as well as integration challenges for the overall process.

2.3.2.1 Feedstock/process interface R&D needs

In the overall context of maximizing fuel production from available biomass, thermochemical gasification approaches will be called upon, because of their inherent robustness, to process low-grade or “non-fermentable” feedstocks. Refinements in dry biomass feeder systems will be required to handle a diversity of feedstocks. These refinements should reduce upfront feed-processing requirements to minimize costs and ideally be capable of delivering feedstocks at less than 15% moisture to the gasifier. If pressurized gasifiers are used, the additional challenges that will be associated with feeding the biomass into pressurized systems must be addressed.

2.3.2.2 Gasification studies R&D needs

Gas-to-liquid fuel synthesis processes are very sensitive to gas quality and composition issues and because gas quality and composition are directly related to the feedstock being gasified, gasification studies will need to determine how feedstock composition affects syngas composition, quality, and efficiency. Additionally, since gasifier technology also has a major effect on gas quality and composition, these feedstock-specific studies need to be correlated with the different gasification technologies.

2.3.2.3 Cleanup and conditioning R&D needs

Techno-economic analysis (40) has shown that removing chemical contaminants such as tar, ammonia, chlorine, sulfur, alkali metals, and particulates has the single greatest effect on the cost of liquid fuel synthesis. To date, gas cleanup and conditioning technologies are unproven in integrated biorefinery applications. Although water quench is an effective approach for removing tars and other particulates from the syngas, it is highly problematic from efficiency and waste disposal perspectives. Therefore, developing catalytic consolidated tar and light hydrocarbon reforming is desirable to eliminate the need for water quench.

An appropriate target for tar and light hydrocarbon reforming is to convert all tars and light hydrocarbons to syngas to a sufficient level as to not require an additional steam methane-reforming unit operation. Specific research needed to accomplish these objectives is as follows:

- Perform tar deactivation/regeneration cycle tests to determine activity profiles to maintain the required long-term tar-reforming catalyst activity

- Perform catalyst studies to determine deactivation kinetics and mechanisms by probing catalyst surfaces to uncover molecular-level details
- Determine optimized catalyst formulations and materials at the pilot scale to demonstrate catalyst performance and lifetime as a function of process conditions and feedstock
- Design catalysts with higher tolerances for sulfur and chlorine poisons to enable further process intensification
- Lower or eliminate the sulfur and chlorine removal cost prior to reforming to achieve further reductions in gas cleanup costs
- Optimize the water gas shift activity of reforming catalysts to reduce or eliminate the need for an additional downstream shift reactor.

2.3.2.4 *Catalytic fuels (mixed-alcohol) synthesis R&D needs*

Although, as stated earlier, there are a number of gas-to-liquid processes capable of producing liquid transportation fuels from biomass syngas, a mixed-alcohol synthesis process is specifically discussed here because of the synergies with the biochemical approach in producing ethanol as the primary product. The commercial success of mixed-alcohol synthesis has been limited by poor selectivity and low product yields. Single-pass yields are on the order of 10% syngas conversion (38.5% carbon monoxide conversion) to alcohols, with methanol typically being the most abundant alcohol produced (41, 42). For mixed-alcohol synthesis to become an economical commercial process, improved catalysts are needed (43). Improvements in mixed-alcohol synthesis catalysts could increase alcohol yields and the selectivity of ethanol production from clean syngas, as well as improve the overall economics of the process through better heat integration and control and fewer syngas recycling loops. Specific research needed to accomplish these objectives is as follows:

- Develop improved mixed-alcohol catalysts that increase the single-pass carbon monoxide conversion from 38.5 to 50% (and potentially higher) and improve the carbon monoxide selectivity to alcohols from 80 to 90%.
- Develop improved mixed-alcohol catalysts with higher activity that require a lower operating pressure (1000 psia compared with 2000 psia) to decrease process-operating costs. The combination of lower syngas pressure for alcohol synthesis and less unconverted syngas to recompress and recycle has the added benefit of lowering the energy requirement for the improved synthesis loop.
- Explore alternative mixed-alcohol synthesis reactors and catalysts. Greatly improved temperature control of the exothermic synthesis reaction has been demonstrated to improve yields and product selectivity. Precise temperature control reactor designs need to be developed for the mixed-alcohol synthesis reaction to improve the yields and economics of the process.

2.3.2.5 *Integration/demonstration needs*

For any sophisticated conversion process, combining individual unit operations into a complete, integrated, systematic process is a challenge. To demonstrate economic competitiveness, individual pilot-scale operations and complete integrated pilot development runs will

be required. A specific challenge is to continue to demonstrate process intensification and higher yields at pilot scale to reduce capital costs.

2.4 Advanced biorefinery

Achieving near-term economic competitiveness with gasoline and starch-based ethanol will enable a viable lignocellulosic ethanol industry. Ethanol from lignocellulosic feedstocks will then be able to join starch-based ethanol in providing a sustainable, renewable resource for the world's transportation needs. However, market analysis (4) indicates that supplying a significant fraction of transportation fuel needs with ethanol for the long-term will require additional technology advancements in all areas of the biorefinery. This is predominately driven by the need to capture higher cost feedstocks to maximize the overall impact of the biorefinery.

Future R&D efforts will need to focus on four complementary approaches. Independently, the approaches will not be sufficient to meet the long-term goals of the biorefinery, but taken collectively they will combine revolutionary scientific breakthroughs with evolutionary process developments to maximize the potential of the biorefinery concept to supply a significant fraction of transportation fuel needs.

Some cost reductions will be achieved by continuous process improvements to near-term technologies. For example, the construction and operation of full-scale biorefineries will highlight opportunities for unit operation optimization and provide operational experience for process optimization and cost reductions. The accumulation of operating experience and engineering data will lead to larger-scale biorefineries, which will further reduce bio-fuels production costs by leveraging economies of scale. These are the evolutionary cost reductions. However, more dramatic cost reductions will be required from scientific breakthroughs for biorefineries to reach their ultimate potential.

Earlier sections of this chapter described technologies for feedstock supply systems, biochemical conversion, and thermochemical conversion to accomplish near-term economic competitiveness. In the future, advances will be made in all three areas, and there will be opportunities for cost savings through integrating biochemical and thermochemical conversion technologies into larger facilities. The four areas of future technology advancement needed to accomplish the ultimate potential of biorefineries are as follows:

- 1 Advanced, large-tonnage feedstock supply systems
- 2 Systems biology to improve biochemical processing
- 3 Selective thermal transformation to improve thermochemical processing
- 4 Technology integration, economies of scale, and evolutionary process optimization.

2.4.1 Advanced, large-tonnage feedstock supply systems

Although it is envisioned that the first Phase III or lignocellulosic biorefineries will capture the low cost or niche feedstocks, eventually a significant fraction of the feedstock potential identified by Perlack and coworkers (3) will need to be captured economically. As the biorefining industry expands, process improvements will drive biorefinery capacities up.

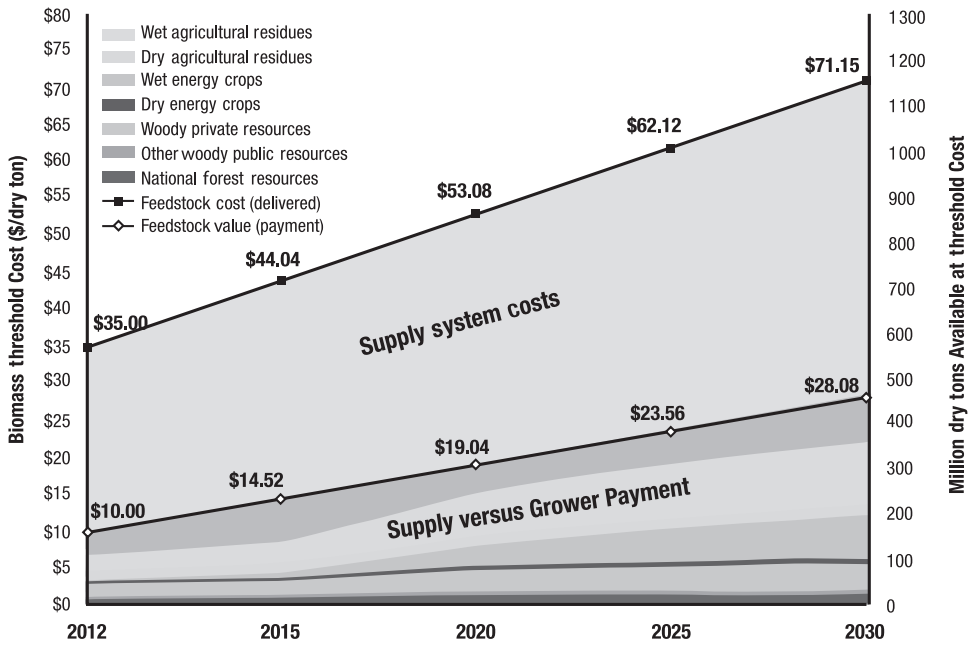


Figure 2.10 Advanced feedstock supply system technologies: value-add feedstock preprocessing.

Therefore, the longer-term feedstock supply R&D challenge is to ensure supply systems do not limit biorefinery size or consume biorefinery profits that could be used to purchase higher-cost feedstocks (see Figure 2.10). Feedstock availability, as a function of payment to the grower (4), has shown that more than 2/3 of the feedstock potential identified by Perlack and coworkers could be made available to the biorefinery at a purchase price up to about \$50/ton. Adding estimated feedstock supply system costs gives a final feedstock cost of about \$70/ton.

An advanced feedstock supply system will be needed to collect the large tonnages of feedstock required for large-scale biorefineries. An efficient interface between producers and the commodity biomass system is important for large-scale feedstock supply technology development. Production, harvesting, and collection systems will be widely varied, based on biomass resources and local practices. Primary research needs include storage, preprocessing, and transportation systems suited to these varied systems.

Developing value-add feedstock preprocessing and blending technologies will provide flexibility in the biomass feedstock supply system and allow suppliers to:

- 1 Reformat/condition different feedstocks into a common format and quality
- 2 Fractionate secondary co-products for local markets
- 3 Produce blended, large-scale commodity biomass.

Value-added preprocessing will help create a market specification for feedstocks, which will help in the transition of biomass to a large-scale commodity and ensure that feedstocks from varied sources can supply a large-scale biorefinery without process upset.

2.4.2 Systems biology to improve biochemical processing

Systems biology research will result in improvements to feedstock and will maximize the recoverable liquid fuel per acre of land and drastically simplify the conversion process. These improvements have the potential to reduce the cost of converting lignocellulosic biomass to ethanol by about 30% for a similar sized 2000 tonnes/day facility (4). Additionally as the advanced state of technologies facilitate larger scale facilities, an additional 40% cost-of-production benefit could be realized for a 10 000 tonnes/day operation. These kinds of cost reductions are typical of conversion technologies as they mature. The oil and corn industries, among others, have seen processing costs drop dramatically over time until feedstock is the predominate cost.

It is envisioned that, through systems biology, the overall conversion process can be simplified, and capital and operating costs can be reduced (see Figure 2.11).

The advanced technology will combine several unit operations and improve the pre-treatment operation. Enzyme production and fermentation will be combined in a single organism. Thus, with enzymes produced during the saccharification and fermentation processes, the three process operations are combined into one. In addition, more robust microorganisms will eliminate the need for hydrolyzate conditioning. These technology improvements will lower the total capital cost (project cost) of a 2000 dry tonnes/day facility by about 22% (44). Further capital cost reductions can be realized as these systems biology advances enable larger scale facilities that take better advantage of economies of scale.

Translational science concepts need be adapted to pursue these advancements. This approach, familiar to the biomedical industry, integrates basic research (or fundamental biological science) with industrial application (such as bioengineering). To meet the long-term potential of the biorefinery, significant fundamental scientific advances beyond what was described for a near-term economic competitive state of technology must be achieved (45). Additionally, it is important that these advances be implemented to realize the significant operating and capital cost reductions, so that the growth of biofuels does not stagnate due to the need to draw in higher cost feedstocks.

2.4.2.1 Fundamental biological science

A full and detailed integration of science and engineering research will be needed for the continued growth of the biorefinery and biofuels. Fundamental R&D in biomass conversion must be targeted to process improvements based on technical barriers. An integrated fundamental and applied research program in biochemical conversion must include advancements in these three critical areas.

2.4.2.1.1 FEEDSTOCK ENGINEERING

- Develop genomics and agronomic/silviculture strategies to maximize the yield and quality of developing energy crops.
- Design and manipulate plant cell wall composition and structure to maximize the yield of fermentable sugars.

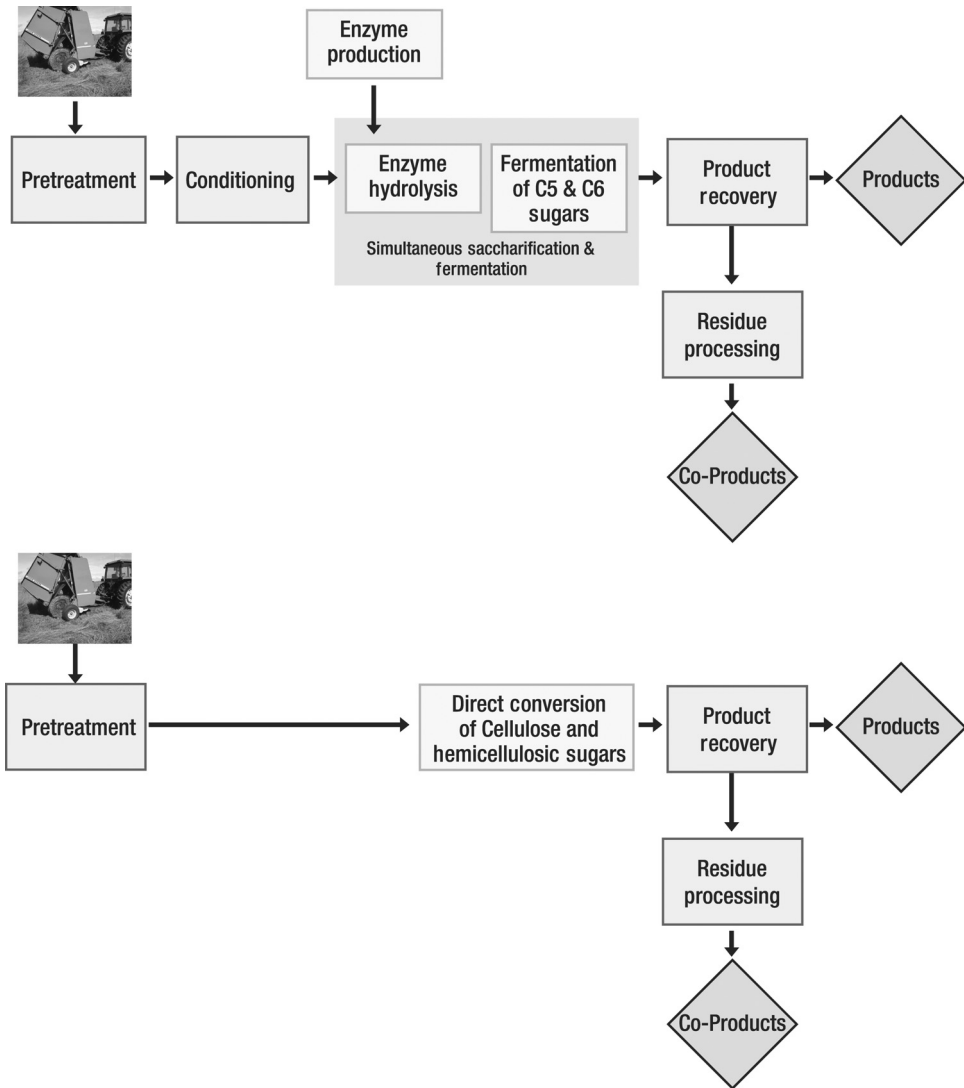


Figure 2.11 Process simplification through systems biology research.

2.4.2.1.2 CELL WALL SACCHARIFICATION

- Analyze glycosyl hydrolase structure/function as it applies to plant cell wall deconstruction.
- Develop improved (engineered) enzymes for advanced biochemical conversion technologies and integrate them with pretreatment chemistries.

2.4.2.1.3 STRAIN DEVELOPMENT

- Apply systems biology and biochemistry to strain improvement to increase the conversion of sugars released during biomass deconstruction to ethanol and products.

- Focus on strains that will produce saccharifying enzymes and ferment the resulting sugars to ethanol.

2.4.2.2 Bioengineering research

The objective of bioengineering research is to acquire new understanding in broad-based aspects of applied process engineering research. Applying commercial enzyme preparation components to various pretreated biomass samples, both within and beyond the scope of established consortia, is also a key component of this research. Work will need to be extended to include comparative pretreatment analysis for multiple feedstocks (e.g., corn stover, switchgrass, and hybrid poplars) and additional pretreatment process impacts (e.g., identifying hydrolyzate conditioning requirements for different pretreatments).

The applied research program required will need to include advances in process application knowledge at two levels. The first will address process-related engineering research that converts new understanding from fundamental research to the biorefinery context. The second will use process-related engineering information to develop industry recommendations regarding process parameters, equipment, and operating conditions.

2.4.3 Selective thermal transformation to improve thermochemical processing

Achieving the near-term economic competitiveness target with gasoline and starch-based ethanol outlined above for biomass gasification – mixed-alcohol synthesis – requires improvements in catalytic tar and light hydrocarbon reforming to increase conversion efficiencies and reduce the capital costs of syngas cleanup and conditioning. However, as is the case for biochemical conversion, significant technological advances beyond this state of technology will be required to realize the ultimate potential of thermochemical conversion approaches for biofuels production. A strategy for accomplishing this involves moving forward with two complementary approaches:

- Pursue scientific achievements to improve yields and efficiencies and maximize process integration opportunities in existing thermochemical processes
- Implement a rigorous research program to investigate fundamental biomass thermochemical conversion to enable alternative processes that will help erase the lines between gasification and pyrolysis as separate technology options.

R&D efforts for breakthrough thermochemical technology should focus on the front-end processes, while the downstream unit operations continue to be optimized. Significant improvements in catalytic gasification will be needed to increase carbon conversion efficiencies to syngas and decrease tar formation. An example of a breakthrough technology would be converting 50% of the methane produced to syngas while simultaneously increasing the throughput of the gasifier by 25%. This improved technology would reduce thermochemical conversion cost by an estimated 38% over the technology listed above (46).

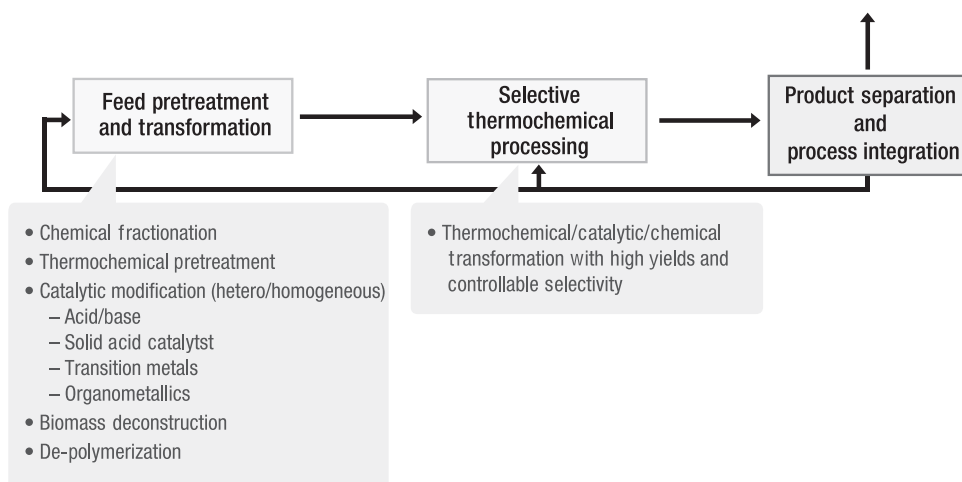


Figure 2.12 Selective thermochemical processing.

Process consolidation is needed to continue lowering capital and operating costs. The block flow diagram in Figure 2.12 illustrates the R&D required to advance thermochemical conversion technology to meet the long-term goals of the biorefinery and biofuels. The following sections describe the research needed to accomplish advanced state of technology for thermochemical conversion shown in Figure 2.12.

2.4.3.1 *Catalytic gasification and pyrolysis*

Since the beginning of coal gasification, catalysts have been sought that would improve carbon conversion to products and increase gasification rates, while minimizing temperature to increase process efficiency. Alkali metals have long demonstrated catalytic activity in steam gasification of solid fuels, and metal-based catalysts – particularly nickel-based materials – are active and effective for hydrocarbon reforming. More R&D along these lines is needed to achieve higher carbon conversions and increased efficiencies in gasification. In the area of pyrolysis, catalytic pyrolysis processes are needed to improve selectivities to more desirable compounds.

2.4.3.2 *Lignin utilization*

Integrating and using lignin residues produced from biochemical-based biorefineries will be key factors in establishing the long-term viability of lignocellulosic biorefineries and maximizing biomass use for fuel production. First-generation biorefineries, call for the lignin to be simply burned to supply the heat and power needs of the biorefinery. Although this is economically viable in the near-term, ultimately because lignin is a complex but lower-value biomass component, it is essential that new technologies increase its value to enhance the competitiveness of integrated biorefineries.

2.4.3.3 Selective thermal transformation of fractionated biomass

A range of alternative conversion options is envisaged through the fractionation of biomass into specific components. A narrower, more uniform biomass fraction opens the possibility of developing thermochemical conversion options with high yields and selectivities.

2.4.4 Technology integration, economies of scale, and evolutionary process optimization

Biochemical and thermochemical conversion technologies can be integrated into a large-scale biorefinery for additional efficiency, yield, and cost improvements. For example, an advanced state-of-technology biorefinery optimized for maximum ethanol production could use an approach as follows: biochemical conversion extracts the carbohydrate portion of the feedstock and then converts it to fermentable sugars and, ultimately, ethanol. The remaining residue, primarily lignin, cannot be fermented, but it is a valuable organic feedstock. By directing this byproduct to a thermochemical process, it can be converted to syngas and, ultimately, ethanol. Integrating these technologies will improve the energy efficiency of the process, lower costs, and produce more ethanol than a standalone biochemical or thermochemical process.

Figure 2.13 depicts the advanced, integrated biochemical and thermochemical alcohol production scheme analyzed. Some of the lignin-rich residue is used to provide steam and electricity to the biochemical process, and the remainder is processed in the thermochemical process. The larger biochemical processes (8000–10 000 tonnes/day) expected in the 2020–2030 time frame will be needed to feed a reasonably sized gasification plant (1500–2000 tonnes/day) with only lignin-rich residues. The scale of the biochemical processing plant is five times larger than that targeted for near-term biorefineries, but the scale of the thermochemical conversion plant is the same.

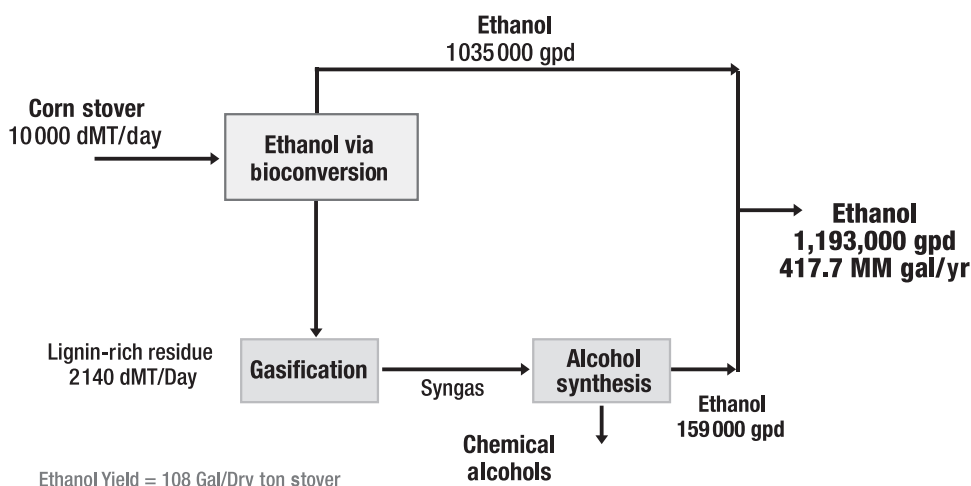


Figure 2.13 Integrated biorefinery gasification scenario with excess lignin converted to ethanol.

This combined process can maximize feedstock handling efficiencies and heat and power integration. Integrated biorefineries can also process feedstocks with both high and low carbohydrate contents. A steady supply of low-carbohydrate feedstock could be fed directly into the thermochemical process, which allows increased size and some benefit to the capital cost. Integrated biochemical–thermochemical biorefineries also capitalize on the process improvements identified in the independent developments of the two technologies.

References

1. Exxon Mobil “The Outlook for Energy – A View to 2030”.
2. US Department of Energy (2003) Energy Information Agency. Available from: http://www.exxonmobil.com/Corporate/energy_outlook.aspx
3. Perlack, R.D., Wright, L.L., Turhollow, A.F., Graham, R.L., Stokes, B.J. & Erbach, D.C. (2005) *Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply*. DOE/GO-102005-2135. Oak Ridge National Laboratory, Oak Ridge, TN.
4. Foust, T.D., Wooley, R., Sheehan, J., Wallace, R., Ibsen, K., Dayton, D., Himmel, M., Ashworth, J., McCormick, R., Melendez, M., Hess, J.R., Kenney, K., Wright, C., Radtke, C., Perlack, R., Mielenz, J., Wang, M., Synder, S. & Wery, T. (2007) *A National Laboratory Market and Technology Assessment of the 30x30 Scenario*. NREL/TP-510-40942. National Renewable Energy Laboratory, Golden, CO.
5. US Department of Energy (2007) Annual Energy Outlook – With Projections to 2030 DOE/EIA-0383. Available from: www.eia.doe.gov/oiaf/aeo
6. Greene, N., Celik, F.E., Dale, B., Jackson, M., Jayawardhana, K., Jin, H., Larson, E.D., Laser, M., Lynd, L., MacKenzie, D., Mark, J., McBride, J., McLaughlin, S. & Saccardi, D. (2005) *Growing Energy: How Biofuels Can Help End America's Oil Dependence*. Natural Resources Defense Council, New York.
7. Kamm, B., Kamm, M., Gruber, P.R. & Kramus, S. (2006) Biorefinery systems – an overview. In: *Biorefineries – Industrial Processes and Products Status Quo and Future Directions* (eds. B. Kamm, P.R. Gruber & M. Kamm). Wiley-VCH, Weinheim, Germany.
8. National Renewable Energy Laboratory (NREL). Available from: <http://www.nrel.gov/biomass/biorefinery.html>
9. Keller, F.A. (1996) Integrated bioprocess development for bioethanol production. In: *Handbook of Bioethanol: Production and Utilization* (ed. C. Wyman). Taylor and Francis Publishers, New York.
10. Kamm, B. & Kamm, M. (2004) Principles of biorefinery. *Applied Microbiological and Biotechnology*, **64**, 137–145.
11. Van Dyne, D.L., Blasé, M.G. & Clements, L.D. (1999) A strategy for returning rural America to long-term full employment using biomass refineries. In: *Perspectives on New Crops and New Uses* (ed. J. Janick), ASHS Press, Alexandria, VA.
12. McAloon, A., Taylor, F., Yee, W., Ibsen, K. & Wooley, R. (October 2000) Determining the cost of producing ethanol from corn starch and lignocellulosic feedstocks, NREL/TP-580-28893.
13. Renewable Fuels Association – Ethanol Biorefinery Locations. Available from: www.ethanolrfa.org/industry/locations
14. Rendleman, C.M. & Shapouri, H. (2007) *New Technologies in Ethanol Production*, USDA, Report Number 842.
15. Ramamorth, R., Kastury, S. & Smith, W.H. (eds.) (2000) *Bioenergy: Vision for the New Millennium*. Science Publishers, Enfield, NH.

16. Huber, G.W., Iborra, S. & Corma, A. (2006) Synthesis of transportation fuels from biomass: Chemistry, catalysts, and engineering. *Chemical Review*, **106**, 4044–4098.
17. Aden, A., Ruth, M., Ibsen, K., Jechura, J., Neeves, K., Sheehan, J., Wallace, B., Montague, L., Slayton, A. & Lukas, J. (2002) *Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis for Corn Stover*. NREL/TP-510-32438. National Renewable Energy Laboratory, Golden, CO.
18. Energy production from biomass (part 2): Conversion technologies. *Bioresource Technology*, **83**(1), May 2002, 47–54.
19. Spath, P.L. & Dayton, D.C. (2003) *Preliminary Screening – Technical and Economic Assessment of Synthesis Gas to Fuels and Chemicals with Emphasis on the Potential for Biomass-Derived Syngas*. NREL/TP-510-34929. National Renewable Energy Laboratory, Golden, CO.
20. Bridgwater, A.V. & Peacocke, G.V.C. (2000) Fast pyrolysis processes for biomass. *Renewable and Sustainable Energy Reviews*, **4**(1), 1–73.
21. Czernik, S. & Bridgwater, A.V. (2004) Overview of applications of biomass fast pyrolysis oil. *Energy and Fuels*, **18**(2), 590–598.
22. Iowa State University – Office of Biomass Program. Available from: www.iastate.edu/~nscentral/news/2007/apr/biofuels.shtml
23. Putsche, V. (1999) Complete process and economic model of syngas fermentation to ethanol, NREL Milestone Report. Available from: <http://devafdc.nrel.gov/bcfdoc/3997.pdf>
24. Huber, G.W., Chheda, J.N., Barrett, C.J. & Dumesic, J.A. (2005) Production of liquid alkanes by aqueous-phase processing of biomass-derived carbohydrates. *Science*, **308**, 1446–1450.
25. Huber, G.W., Cortright, R.D. & Dumesic, J.A. (2004) Renewable alkanes by aqueous-phase reforming of biomass-derived oxygenates. *Angewandte Chemie*, International Edition, **43**(12), 1549–1551.
26. Huber, G.W., Shabaker, J.W. & Dumesic, J.A. (2003) Raney Ni–Sn catalyst for H₂ production from biomass-derived hydrocarbons *Science*, **300**(56289), 2075–2077.
27. Magrini-Bair, K.A., Czernik, S., French, F., Parent, Y.O., Chornet, E., Dayton, D.C., Feik, C. & Bain, R. (2007) Fluidizable reforming catalyst development for conditioning biomass-derived syngas. *Applied Catalysis A: General*, **318**, 199–206.
28. Stoppert, J. (April 2005) *Industrial Bio-Products: Where Performance Comes Naturally*. BIO World Congress Meeting. Orlando, FL.
29. Cushman, J.H., Easterly, J.L., Erbach, D.C., Foust, T.D., Graham, R., Hess, J.R., Hettenhaus, J.R., Hoskinson, R.L., Perlack, R.D., Sheehan, J.J., Sokhansanj, S., Tagore, S., Thompson, D.N., Turholow, A. & Wright, L.L. (2003) *Roadmap for Agricultural Biomass Feedstock Supply in the United States*. DOE/NE-ID-11129 Rev. 1. Idaho National Engineering and Environmental Laboratory, Idaho Falls, ID.
30. Hakkila, P. (2004) *Developing Technology for Large-Scale Production of Forest Chips: Wood Energy Technology Programme 1999–2003*. Tekes, Helsinki, Finland.
31. Wright, J. (1987) Fuel ethanol technology evaluation. In: *Biofuels and Municipal Waste Technology Research Program Summary FY 1986*. DOE/CH/1093-6, National Technical Information Service Report DE87001140. Solar Energy Research Institute, Golden, CO.
32. Sheehan, J. & Riley, C. (2001b) *Annual Bioethanol Outlook: FY 2001*. Internal NREL-DOE report. National Renewable Energy Laboratory, Golden, CO.
33. Jechura, J. (October 11, 2005) Sugar platform post enzyme-subcontract case. NREL technical memo.
34. Selig, M.J., Knoshaug, E.P., Decker, S.R., Baker, J.O., Himmel, M.E. & Adney, W.S. (Heterologous expression of *Aspergillus niger* B-D-Xylosidase (XlnD): Characterization on lignocellulosic substrates. *Applied Biochemistry and Biotechnology*. Published online November 8, 2007.
35. Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y.Y., Holtzapple, M. & Ladisch, M. (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology*, **96**, 673–686.

36. Teymouri, F., Laureano-Perez, L., Alizadeh, H. & Dale, B.D. (2005) Optimization of the ammonia fiber explosion (AFEX) treatment parameters for enzymatic hydrolysis of corn stover. *Bioresource Technology*, **96**, 2014–2018.
37. Terter, S.A., Xu, F., Nedwin, G.E. & Cherry, J.R. (2006) Enzymes for biorefineries. In: *Biorefineries – Industrial Processes and Products Status Quo and Future Directions*, Vol. 1 (eds. B. Kamm, P.R. Gruber & M. Kamm). Verlag Gmbh and Co., Wiley-VCH, Weinheim, Germany.
38. NREL, Genencor press release. Available from: http://www.genencor.com/cms/connect/genencor/media_relations/news/archive/2004/gen_211004_en.htm
39. Phillips, S., Aden, A., Jechura, J., Dayton, D. & Eggeman, T. (2007) Thermochemical ethanol via indirect gasification and mixed alcohol synthesis of lignocellulosic biomass. NREL Technical Report, NREL/TP-510-41168.
40. Aden, A. & Spath, P.L. (September 2005) The potential of thermochemical ethanol via mixed alcohols production. Milestone completion report.
41. Wender, I. (1996) Reactions of synthesis gas. *Fuel Processing Technology*, **48**(3), 1041–1048.
42. Herman, R.G. (2000) Advances in catalytic synthesis and utilization of higher alcohols. *Catalysis Today*, **55**(3), 233–245.
43. Fierro, J.L.G. (1993) Catalysis in C1 chemistry: Future and prospect. *Catalysis Letters*, **22**(1–2), 67–91.
44. Jechura, J. (June 27, 2006) Integrated biochemical and thermochemical biorefinery scenarios involving mixed alcohol production. NREL technical memo.
45. Breaking the Biological Barriers to Cellulosic Ethanol – A Joint Research Agenda – A Research Roadmap Resulting from the Biomass to Biofuels Workshop, June 2006, DOE/SC-00956.
46. Spath, P., Aden, A., Eggeman, T., Ringer, M., Wallace, B. & Jechura, J. (2005) Biomass syngas to hydrogen production design report. NREL Report.

Chapter 3

Anatomy and Ultrastructure of Maize Cell Walls: An Example of Energy Plants

Shi-You Ding and Michael E. Himmel

3.1 Introduction

Lignocellulosic biomass has long been recognized as a potential sustainable source of mixed sugars for fermentation to fuels and other bio-based products. However, the chemical and enzymatic conversion processes developed during the past 80 years are inefficient and expensive. The inefficiency of these processes is, in part, due to the lack of knowledge about the structure of biomass itself; the plant cell wall is indeed a complex nanocomposite material at the molecular and nanoscales. Current processing strategies have been derived empirically, with little knowledge of the nanostructure of the feedstocks, and even less information about the molecular processes involved in biomass conversion. Substantial progress toward the cost-effective conversion of biomass to fuels is contingent upon fundamental breakthroughs in our current understanding of the chemical and structural properties that have evolved in biomass which prevent its disassembly, collectively known as “biomass recalcitrance.”

This chapter is not a strict review of plant anatomy. It deals only with those aspects of plant structure that are believed important to the availability and digestion of cell walls and the breakdown of biomass materials to fermentable sugars through chemical and biological processes. The anatomy and ultrastructure of plant cell walls will be reviewed and emphasis will be given to recent progress made toward gaining an understanding of cell wall biosynthesis as well as characterization of cell walls at the molecular level using atomic force microscopy (AFM) and fluorescence labeling techniques. Future work and new techniques needed for characterization of the molecular architecture of the plant cell walls are also discussed. In this context, plant cell walls from maize (*Zea mays* L.) stem are used as a model to cover what is currently known about cell wall structures related to biomass recalcitrance and subsequent conversion to biofuels.

3.2 Cell wall anatomy

Maize (*Zea mays* L.) is a monocotyledonous plant whose anatomical structure has been well studied. It resembles other grasses in the arrangement of tissues in the stem, leaf, and

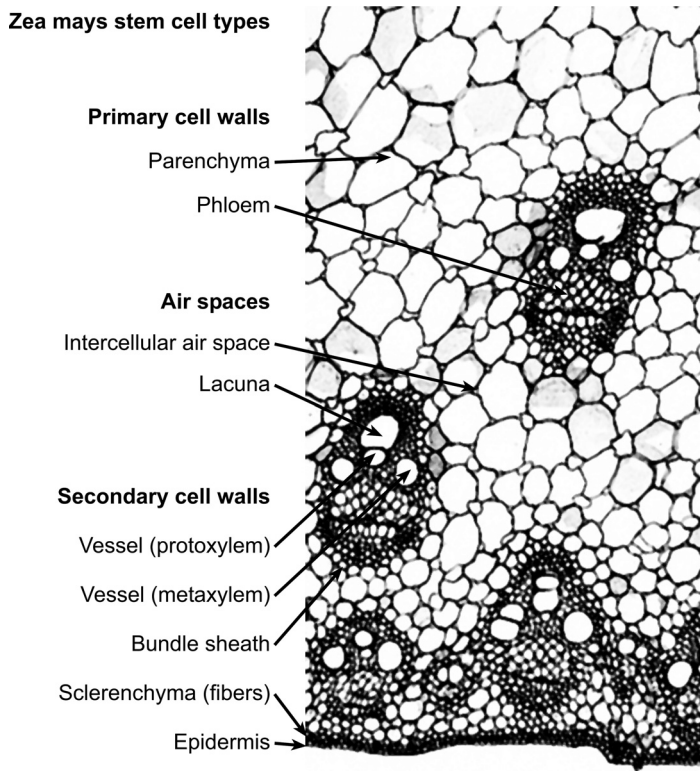


Figure 3.1 Transverse section of maize stem showing cell types.

root. The stems of monocots generally have epidermis, scattered vascular bundles, and parenchyma pith (Figure 3.1).

3.2.1 Plant tissues

Plant stem tissues can be generally categorized as dermal, fundamental, and vascular tissues. The epidermis comprises the surface protection layer, which includes three types of cells: epidermal cells, which have a fairly thick, cuticle-like wall, and the guard and subsidiary cells, which are specialized cells that form the stomata.

The fundamental tissue of maize stem contains three types of cells. Sclerenchyma cells are found immediately beneath the epidermis. Typically, there are 1–3 layers of collenchyma cells, which are non-lignified, and lie below the epidermis. These cells are elongated axially with irregularly thickened walls. There are usually no sclerenchyma cells directly under the stomata. The parenchyma is the most numerous cell type in maize, forming the bulk of the stem. Parenchyma cells often have thin, non-lignified primary walls. The vascular bundles are surrounded by 2–11 layers of bundle sheath (fiber). The fibers are slender elongated cells

with multi-layered secondary walls. As the stem matures, the cells in the outer part of this zone become lignified.

Maize stem has scattered vascular bundles that are distributed throughout section, which is the common arrangement in monocots. These vascular bundles are scattered rather evenly through the stem pith but more numerous near the periphery. Each vascular bundle contains xylem, phloem, and other types of cells. Xylem, which is the tissue that conducts water through the plant, is composed of several different cell types. Protoxylem consists of one or two medium-diameter vessels and surrounding parenchyma. In some bundles, a cavity will be observed in the protoxylem, sometimes referred to as a protoxylem canal or protoxylem lacuna, which represents the position previously occupied by the first-formed protoxylem elements and is essentially an empty space. Metaxylem consists of wide vessels with a few narrow tracheids between them and also some surrounding non-lignified parenchyma. Tracheids are very elongated cells with bordered pit-pairs present along the walls of two adjacent cells. They differ from vessel elements in the imperforate end walls. Phloem, which is the tissue that moves sugars and other products through the plant, is also composed of several specialized cell types. Sieve tubes (wide in cross section) and companion cells (narrow and often darkly stained) form a regular crisscrossed pattern, which is typical of monocots. Most of the phloem is the latter-differentiated metaphloem, but one may also see the remnants of protophloem occurring as an irregular green line toward the outer face of the phloem and beneath the bundle sheath cells.

3.2.1.1 Primary and secondary cell walls

Plant cell walls are complex and dynamic structures composed of a large degree of cross-linked polysaccharide networks, glycosylated proteins, and lignins. Primary cell walls are formed during early cell growth. The primary cell walls vary in thickness among different cell types and are usually not lignified. However, in older tissue, parenchyma cell walls can show various degrees of lignification. The secondary cell wall is deposited when cell growth has ceased and they are often highly lignified. The tracheary elements (tracheids, vessel elements) and fibers are cells that develop rather thick secondary cell walls. Primary cell walls often become lignified when secondary wall formation begins. The cell wall is generally comprised of long cellulose microfibrils interconnected by hemicellulose polysaccharides, such as xylan and xyloglucan. These amorphous hemicelluloses are generally hydrophilic, thus retaining water in the presence of the hydrophobic and highly crystalline cellulose chains and fibers. Additional polymers, such as pectins and lignins, fill in spaces in structures, such as secondary cell walls and the middle lamella.

3.2.1.2 Pore system of cell walls

In mature tissues, there is large volume of intercellular spaces arising by splitting of adjunct cells, or by crushing of entire cells such as lacuna in vascular bundle resulting from the breakdown of protoxylem tissue. The intercellular spaces are functioned as air space or containers for various secreted materials in a living plant.

Plant cell walls and intercellular spaces are defined as the apoplast of the plant body and serve as mechanical support. Cell walls have important effects on the physiological activities of plant tissues, such as absorption, transpiration, translocation, and secretion. These

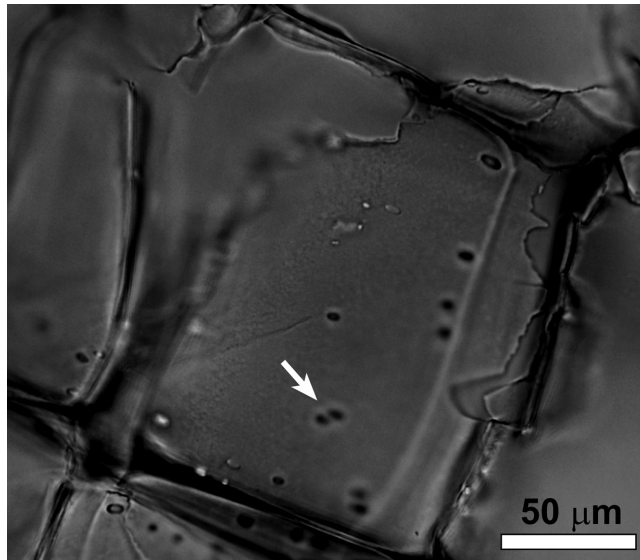


Figure 3.2 Parenchyma cell wall from maize stem pith. White arrow shows the primary pits.

functions are performed through the anatomical features of plant cell walls, particularly the pore systems. Pits, primary pit-fields, and plasmodesmata are commonly observed characteristics of cell walls. Pits are essentially depressions of walls. In the secondary cell wall, two adjunct cells often form opposing pits together called the pit-pair. In primary cell walls, these depressions are also called the primary pit-field, which constitutes a zone of a number of traverse “canals” penetrating the wall, called plasmodesmata. Plant cytoplasm (the symplast) is contiguous and connected by these plasmodesmata that allow the communication between neighboring cells. In mature tissues, secondary wall material may be deposited to form primary pits, which are small in parenchyma tissue (2–5 μm in diameter, see Figures 3.2 and 3.3), whereas in thickened xylem tissues, deposition of thick secondary wall often forms large pits with characteristic patterns (tens of microns in diameter).

3.3 Cell wall biosynthesis and molecular structure

Cell walls from higher plants are primarily composed of polysaccharides (i.e., cellulose, hemicelluloses, and pectins), lignins, glycoproteins, and small amounts of minerals and other polymers. These polysaccharides exist in various forms as crystalline and sub-crystalline celluloses, non-crystalline hemicelluloses, and pectins. The hemicelluloses are closely associated with the surface of the rigid cellulose elementary fibril, forming a microfibril network (1). Pectins are cross-linked polysaccharides that “glue” the cell wall components together. Upon synthesis, these polymers form nanometer scale composites (i.e., microfibrils and matrices) as a result of temporally and spatially controlled processes which occur during plant growth and development. It is still unknown how these polymeric constituents self-assemble to form the rigid and dynamic entity embodied in the cell wall. In our current understanding

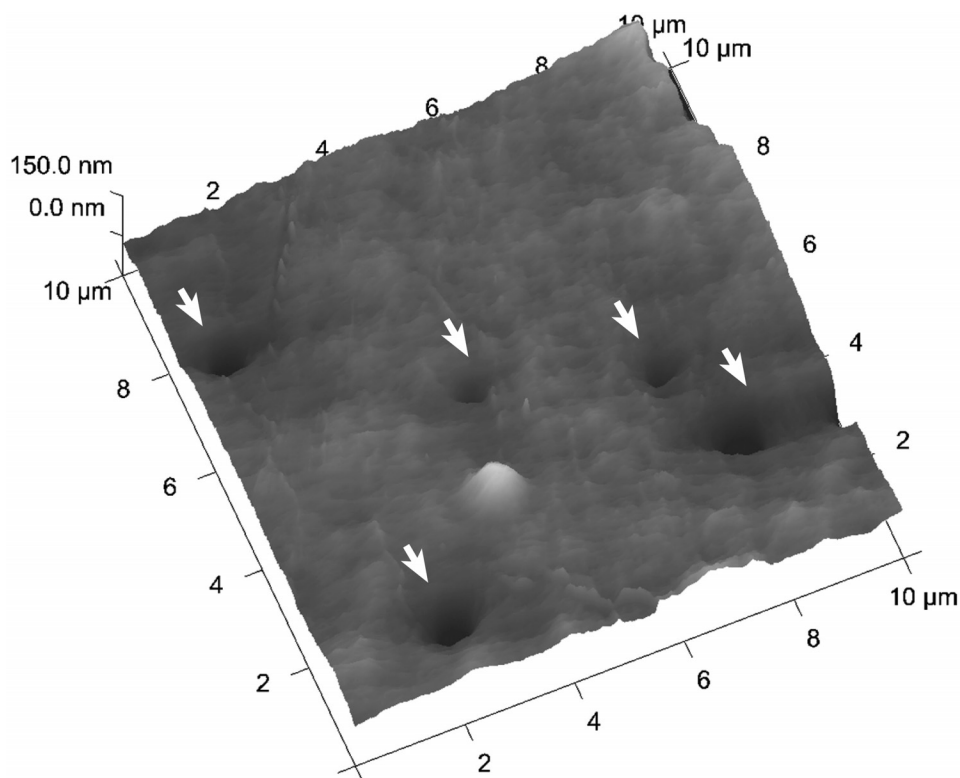


Figure 3.3 Atomic force micrograph mapping the parenchyma cell wall surface. White arrows show the primary pits.

of primary cell wall structures, these associated and/or cross-linked polysaccharides form a “polymer liquid crystal” system (2).

3.3.1 Biosynthesis

Most recent achievements in plant cell wall biosynthesis research resulted from the analysis of *Arabidopsis thaliana* cell wall phenotype mutants and the availability of genome sequence data (see detailed review in Chapter 5) (3, 4). The cellulose synthase complex (CesS), known as rosettes in higher plants, was first observed using electron microscopy and the freeze-fracture sample preparation technique. Rosettes appear in hexagonal geometry with a honeycomb pattern arrayed in the plasma membrane (5, 6). Rosettes are believed to be responsible for the synthesis of elementary fibrils in most current plant cell wall biosynthesis models (7). More recently, mutant analyses [reviewed by Doblin and coworkers (4)] and immunolabeling (8) have confirmed that these rosettes are composed of cellulose synthase (CesA) proteins, and that at least three types of CesA isoforms ($\alpha 1$, $\alpha 2$, and β) are required

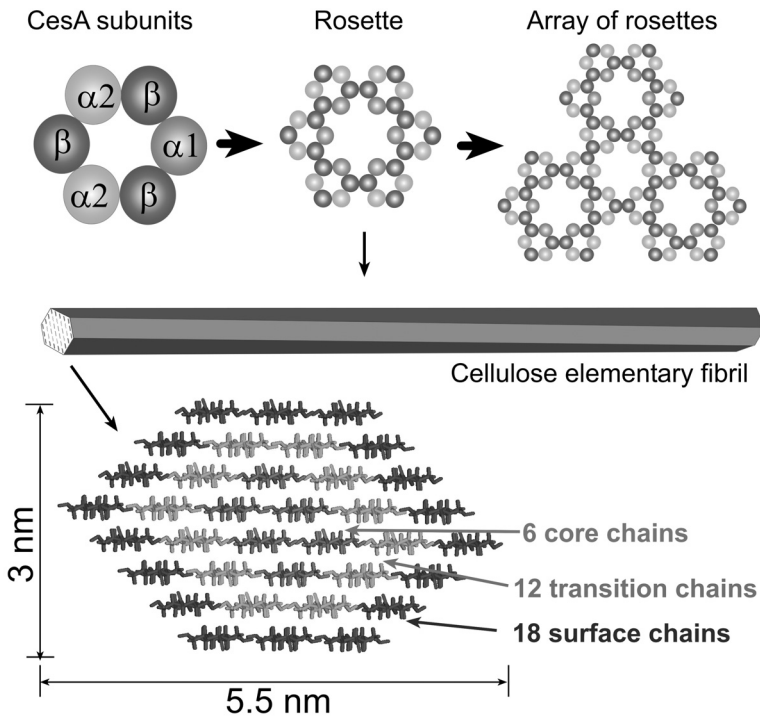


Figure 3.4 Model of plant cell wall cellulose elementary fibril and its synthesis. In this model, at least three types of cellulose synthases (CesA subunits, $\alpha 1$, $\alpha 2$, and β) are needed to spontaneously assemble the rosettes that composed of 6×6 CesA enzymes synthesizing 36-chain cellulose elementary fibril. The rosettes may also form arrays in the cell membrane, in this case, a number of rosettes synthesize a bundle of elementary fibril, the macrofibril. The estimated dimensions of elementary fibril are 3×5.5 nm that agrees with direct measurement using atomic force microscopy (see also Figure 3.9). The depiction of the glucan chains is based generally on an X-ray structure of cellulose I β . It has been proposed that the cellulose elementary fibril may contain three groups of glucan chains: in group C1 (red) there are 6 crystalline chains; in group C2 (green) there are 12 sub-crystalline chains with a small degree of disorder; and in group C3 (blue) there are 18 surface chains that are sub-crystalline with a large degree of disorder. (Modified from Ding and Himmel, 2006; Himmel *et al.*, 2007) (Reproduced in color as Plate 1.)

for the spontaneous assembly of single rosettes. The next question is – how many CesA proteins are assembled into single rosettes?

A model of rosettes has been recently proposed (9) (see Figure 3.4) in which three types of interactions are needed for the spontaneous assembly of rosettes in the plasma membrane, these are β – β , $\alpha 1$ – β , and $\alpha 2$ – β . The β – β interaction facilitates the assembly of six subunits into rosettes, as well as the array of rosettes in plasma membrane. The six subunits of rosettes are identical and each rosette is composed of one molecule of $\alpha 1$, two molecules of $\alpha 2$, and three molecules of β . This simplified rosette model seems probable because of the proposed $\alpha 1$ positioning and the fewer types of interactions required for the rosette assembly in plasma membrane. There is no direct biochemical evidence yet to confirm how the rosettes are assembled, probably the zinc finger domains of CesA proteins, plasma

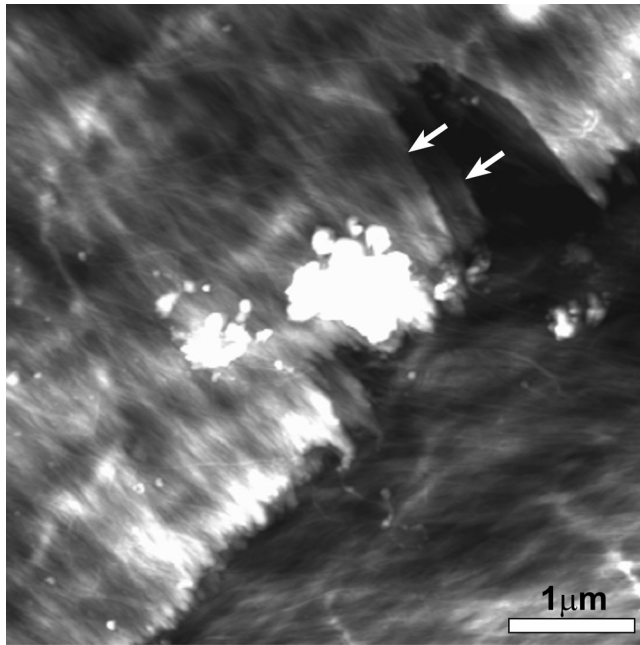


Figure 3.6 Atomic force micrograph shows cell wall layers in a broken edge of maize parenchyma wall.

enzymes. This enzyme complex thus produces 36β -1,4-glucan chains, which simultaneously coalesce to form the cellulose elementary fibril. Hemicelluloses are synthesized in Golgi apparatus and secreted to cell wall. The hemicellulose particle interacts with the surface of newly synthesized cellulose elementary fibrils that usually form a ribbon-like bundle (the macrofibril) (9). During cell wall expansion, the macrofibril is thought to split into single elementary fibrils, the hemicellulose particles unfold and layer upon the microfibrils where some fraction of these polysaccharides start to coat the cellulose surface through numerous hydrogen bonds.

3.3.2 Cell wall lamellae

Cell walls are deposited by layers upon synthesis. Generally, primary cell walls are synthesized when cells grow. Secondary cell walls are deposited when cell growth has ceased. Some cells possess only primary wall, such as the parenchyma cells; however, secondary deposition may occur on most of cell walls when cells age. For example, thin layers of secondary deposition are commonly observed in mature parenchyma (Figure 3.6). Also, there is only one microfibril sheet in each lamella. These thin secondary wall lamellae are measured approximately 10 nm, which appear to contain only one layer of parallel-arranged microfibril. Microfibrils are rotated approximately 50° with respect to each lamella. Secondary cell walls commonly consist of three anatomical layers: the outer (S1), middle (S2), and inner (S3) layers. The thickness of each layer varies in different cell types and tissues. The S2 layer is often thickest, and sometimes contains sub-layers. Another important structure in cells with

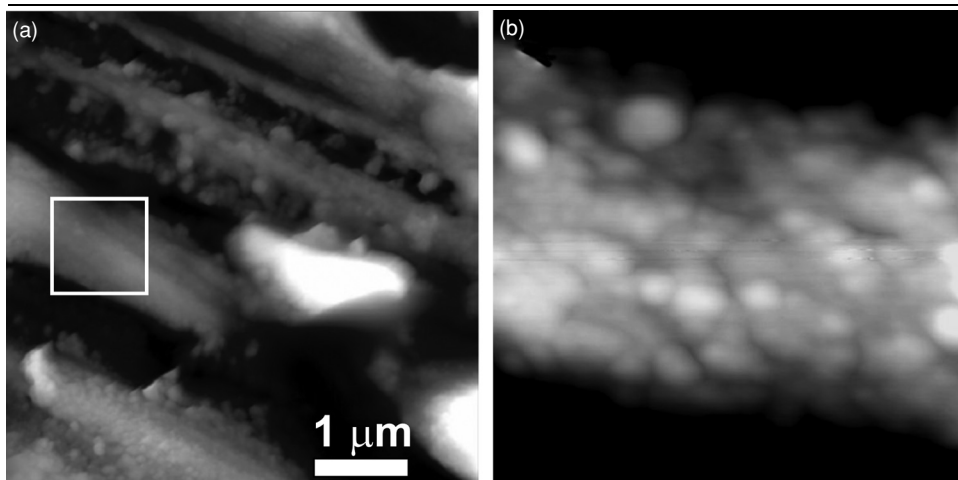


Figure 3.7 Atomic force micrographs show the surface structure of a trichid cell. (a) Secondary cell wall is covered by granules and amorphous structures. (b) Zoom in structures shows the squared area in (a).

thick secondary wall (e.g., vessel cells) is the warty layer in the inner surface, this layer is comprised of granules and amorphous structures (Figure 3.7) that restrict access by rumen microorganisms (12).

3.3.3 *The macrofibril and elementary fibril*

Using AFM, we recently observed composite structures appearing on the inner surface of parenchyma cell walls. These structures appeared to be bundles of fibrils, termed the macrofibril (9). The size of the macrofibril varies from 50 to 250 nm in diameter. Figure 3.8

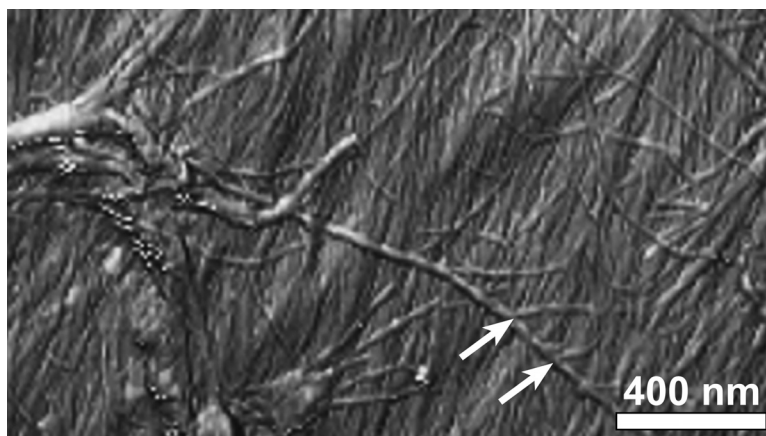


Figure 3.8 Atomic force micrograph of the surface structure of maize parenchyma shows the macrofibril branching at the end.

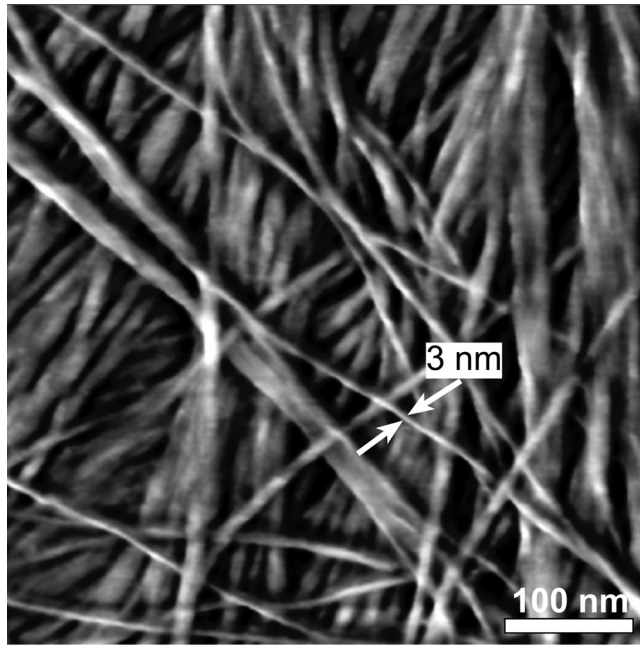


Figure 3.9 High-resolution atomic force micrograph of maize parenchyma wall shows the detailed surface structure of microfibrils. The diameter of individual microfibril is only about 3–5 nm.

shows a large macrofibril “branching” or untangling at the end to form a set of smaller parallel fibers which appears to be the microfibrils. The macrofibril is also observed in our recent study on fresh parenchyma cells. The morphology of fibers in the macrofibril appears to be faceted, which differs from that of the microfibril (Ding, unpublished data).

The elementary fibril is the native form of the plant cell wall cellulose. Direct characterization of the molecular structure of the elementary fibril has not been reported. Figure 3.4 is the recent model we proposed (9). In this model, the 36 cellulose chains are arranged in an I β -like structure, forming a hexagonal geometry in cross section with approximate dimensions of 3×5.5 nm. For this model, there are three distinct layers of chains: 6 group-C1 core chains, 12 group-C2 transition chains, and 18 group-C3 surface chains. The elementary fibrils are considered to be highly crystalline. However, due to the fact that half of the cellulose chains are surface chains and their interactions with the cell wall matrix polymers, elementary fibrils are expected to have some structural disorders.

3.3.4 The microfibril

It is thought that the microfibril contains a single elementary fibril surrounded by hemicelluloses (see Figure 3.9). The dimensions of microfibrils determined by electron microscopy, and recently by AFM, range from 2 to 3 nm to 20 to 50 nm (13). This apparent size variation of the microfibrils may be due to the different plant species and tissue types, but is

more likely due to the sampling errors and inaccurate measurements made using imaging techniques with limited resolution. Current biochemical evidence has shown similar biosynthesis mechanisms (e.g., highly gene homology) for all higher plants (14). There is no evidence to suggest that the elementary fibril is synthesized differently among higher plant species. In contrast, the microfibrils are dynamic structures that could be modified post-formation depending on their interaction with the matrix polymers. The diameter of the microfibrils could be different for different tissues, different cell wall types, or even for different cell wall lamellae.

3.3.5 Cellulose

Cellulose is often considered to be one of the most abundant biopolymers on earth. Although the composition of cellulose, the β -1,4 linked linear glucose polymer (glucan), is relatively simple compared to other plant cell wall polysaccharides and the physical structure of cellulose is complicated. Celluloses can be crystalline, sub-crystalline, and even amorphous, depending on their tissue source in native plant, or the way that cellulose is isolated. The structural integrity of cellulose is believed to be one of the major causes of resistance to chemical and enzymatic hydrolysis.

Until recently, the crystal structures of native cellulose (cellulose I) were elucidated using algae *Glaucocystis* (I α , triclinic) (15) and the tunicate *Halocynthia roretzi* (I β , monoclinic) (16) celluloses. It has been suggested that the I α and I β allomorphs naturally coexist in various proportions in different organisms (17). These celluloses obtained from algae or tunicate; however, do not necessarily represent the cellulose in plant cell walls. The cellulose crystallite in higher plants is reported to be much smaller (3–5 nm in diameter) (9, 17), whereas they can be approximately 20 nm in diameter in some algae. Plant cell wall cellulose is embedded in a complex polymer matrix forming the microfibrils. Indeed, the interactions between cellulose and other polysaccharides are ubiquitous. The plant cell walls are dynamic and their compositions and properties may differ for different tissues, cell types, as well as different developmental stages. While the molecular structure of plant cell wall cellulose remains unclear, some scientists believe that cellulose in higher plant cell walls exists primarily in the I β form with a small proportion of I α form (13, 18–20). Other researchers have suggested that there is only the I β form, with some disorder on the crystallite surface, based on solid-state ^{13}C NMR spectroscopy studies (21). The challenges for characterizing plant cell wall celluloses stem not only from the limited resolution of available measurement techniques, but also from the processes commonly used to isolate celluloses, as well as sample preparation methods. Typically, sequential extraction processes using acid and alkaline incubations are used, sometimes at high temperature. Such extensive sample processing can result in fiber aggregation (22). The isolated “microfibrils” are thought to be cellulose crystallites surrounded by a small pocket of partially hydrolyzed non-crystalline polysaccharides (13, 23). The question remains, how does this extensively treated “microfibril” relate to its native form in the plant cell wall? The plant cell wall cellulose crystallites can be summarized as being too small for traditional microscopy (3–5 nm in cross section) and containing a large proportion of disordered surface glucan chains (20, 24, 25). The detailed molecular structure of plant cell wall cellulose remains unknown.

3.3.6 Matrix polymers

Matrix polymers in plant cell walls are polysaccharides (hemicelluloses and pectins) and lignins. In contrast to cellulose, the matrix polysaccharides are synthesized in the Golgi apparatus, and deposited into the cell wall network while cellulose is synthesized (26). These polymers form a matrix network directly or indirectly associated with the surface of cellulose elementary fibrils. Matrix polysaccharides are non-crystalline structures and vary in glycosidic linkages, branching chemistry, and sugar residues. Lignins are polymers of monolignols (phenylpropane derivatives) providing mechanical strength and water resistance for the plant cell wall, as well as resistance of microbial attack (27–29). Lignin deposition is the final differentiation stage of plant cells that have thick secondary cell walls and proceeds via several phases, starting at the cell corners in the region of the middle lamella and the primary wall after the S1 layer formation has initiated. When the formation of the polysaccharide matrix of the S2 layer is completed, lignification proceeds through the secondary wall. The bulk of lignins are deposited after cellulose and hemicellulose have been deposited in the S3 layer. This is why lignin concentration is higher in the middle lamella and cell corners than in the S2 secondary wall (29–31).

3.4 Advanced approaches for characterizing cell wall structure

Whereas plant cell walls have been measured by electron microscopy (EM) (13, 32–36), the drawback of EM techniques lies in the sample preparation and imaging processes commonly used, which involve chemical extraction, dehydration, and embedding. Because biomass samples go through both chemical pretreatment and enzymatic hydrolysis, methods must be developed to minimize the disruption of sample preparation, and conduct imaging under physiological conditions. Recently developed imaging methods, such as AFM, nonlinear optical microscopy, and single molecule methods, are capable of imaging the cell wall at nanoscale resolution without extensive sample preparation. These techniques can also be applied to directly map the macromolecular structures of cell walls, and to visualize their degrading enzymes simultaneously.

3.4.1 Atomic force microscopy

AFM measures attractive or repulsive forces between a probe or “tip” and the sample. The height imaging (*z*-axis) measures the topography and the phase imaging detects variations in composition, adhesion, friction, viscoelasticity, and perhaps other properties. Phase imaging is particularly useful for mapping variations in sample properties at very high resolution, often with superior image detail. AFM is also able to image samples under fluids that could directly visualize biomacromolecules under physiological conditions (37). AFM has been used increasingly to map cell wall surface structure. Cell walls of various plant species, both from the Monocotyledonae and the Dicotyledonae, have been imaged using AFM (9, 38–42). Early AFM studies of cell walls mostly relied on extensive sample preparation processes using grinding and/or chemical extraction (37–40, 43), which could disrupt

the native structure of cell wall fibrils (34). Recently, fresh cell walls from intact tissues have been imaged in water or in a partially hydrated state resulting in varied observations. In one case (42), celery epidermal peels were imaged under water with sequential dehydration conditions (e.g., using various concentrations of ethanol). The microfibrils were found to be smaller, uniformly distributed, and highly parallel in the never-dried specimen. However, the microfibrils were found to be distinctly larger and more disorganized after dehydration treatment with ethanol or air. These investigators then concluded that dehydration processes could significantly affect the structure and arrangement of primary cell wall microfibrils. Our previous study (9) revealed an accurate measurement of maize cell wall dimensions with nanoscale resolution. Figure 3.9 shows a high-resolution image of maize parenchyma cell wall taken by AFM phase imaging. For example, a single microfibril can be measured in maize primary walls and it is approximately 3–5 nm in diameter and tens to hundreds of microns in length (supporting our previously proposed model).

3.4.2 Biophotonics and nonlinear microscopy

The combination of femtosecond lasers and the high numerical aperture optics found in microscopy makes it possible to create high intensities (100 GW/cm^2) with extremely modest energies ($\sim 100 \text{ pJ}$). This high intensity results in inducing a dynamic, nonlinear polarization in virtually any media located within the focal volume of the microscope objective. This nonlinear, time-varying polarization response acts as a driving force in the wave equation that can result in new source terms. These new sources can be used to create image contrast. Because they scale nonlinearly with the excitation intensity they are strictly confined to the focal volume (no out-of-focus contributions), and in essence are naturally confocal. The net result – nonlinear microscopy – is a high-resolution (sub-micrometer lateral resolution, micrometer axial resolution), three-dimensional imaging modality capable of effectively probing material structure and function. While these intensities may seem extreme, the combination of modest energy (44) and infrared wavelengths actually results in a relatively benign excitation source. In comparison to continuous wave excitation at UV or near-UV wavelengths, delicate systems are minimally perturbed under femtosecond laser excitation.

The recently developed nonlinear microscopy has been applied to imaging biological systems, such as nonlinear signals of second (SHG) and third harmonic generation (THG) (45), coherent anti-stokes Raman spectroscopy (CARS) (46, 47). These techniques combine spectroscopy (chemical) and microscopy (spatial) approaches, which have particular potential in characterizing plant cell wall structures and their bioconversion processes.

3.4.3 Single molecule methods

Advances in fluorescence labeling techniques suited to biological applications have resulted in widespread adoption of the total internal reflection fluorescence (TIRF) technique for biophysical studies (48, 49). With the recent development of photo-activated fluorescence

proteins (PA-FP) (50), a new paradigm in single molecule imaging has developed. By sequentially imaging sparse subsets of single molecules, and localizing their centroids with molecular precision, composite optical images can be constructed with up to two orders of magnitude higher spatial resolution than with conventional methods (51, 52). This technique relies on photo-activation of the PA-FP, followed by photo-bleaching or photo-switching, such that only a sparse subset of molecular tags are excited in a given time window. If molecular fluorophores are strategically attached to relevant cellular structures, structural and chemical information of the cell and intracellular constituents may be obtained with nanometer resolution. A catalog of PA-FPs has now been developed for use in fluorescence imaging of living systems (51, 53–58).

Theoretically, any protein molecule could be expressed as a fusion protein with a fluorescence protein (FP) then imaged by TIRF. While TIRF studies are commonplace in the life sciences and nanotechnology, this technique has only recently been adapted to the study of the structure of the plant cell wall (59). In order to specifically label the ultrastructure of the cell wall, molecular probes recognizing cell wall macromolecules have been recently reported, including monoclonal antibodies against polysaccharides (60–64) and lignin (65–67), and CBMs recognizing cell wall polysaccharides (59, 68, 69). Our previous study demonstrated, for example, that the innate binding specificity of different CBMs offers a versatile approach for mapping the chemistry and structure of surfaces that contain complex carbohydrates (59). In nature, the CBM serves as an attachment device for “harnessing” the glycoside hydrolases to their target substrates (70, 71). Several hundred putative CBMs have been identified to date and these proteins have been grouped into 43 families using amino acid sequence similarity algorithms (<http://afmb.cnrs-mrs.fr/CAZY/index.html>). The structures and ligand specificity of many CBMs have also been determined experimentally (71). Among these CBMs, one type, termed surface-binding CBMs, bind specifically to the planar surfaces (1,1,0 and 1,-1,0) of crystalline cellulose I α (72). We have used genetic engineering methods to produce labeled CBMs, for example C α CBM3-GFP is a surface-binding CBM cloned from *C. thermocellum* and tagged with green-fluorescence-protein (GFP). C α CBM6-RFP is a polysaccharide-binding CBM cloned from *C. thermocellum* and tagged with red-fluorescence-protein (RFP). Figure 3.10 shows the TIRF image where green signal highlights the microfibril network structure and red signal shows the location of the cell wall matrix. In our previous study, we confirmed that C α CBM3 binds to crystalline cellulose and thus we believe that the microfibril network contains highly crystalline cellulose (59).

3.4.4 Computer simulations

Because of the complexity of plant cell wall structure in terms of its components and organization, it is difficult to decipher cell wall imaging results without knowledge of its molecular and electronic structures. Even though recent experiments using synchrotron X-ray and neutron diffraction have elucidated the crystal structures of cellulose I α and I β from algae and tunicate respectively (15, 16), the cellulose structures in plant cell walls remain largely unclear. However, it is known that the cross-sectional diameter of an elementary cellulose fibril is only about 3–5 nm in size (9). It has been further proposed (9) that the cellulose elementary fibril consists of 6, 12, and 18 glucan chains in the center, middle, and interface respectively, with increasing disorder. Only the six center chains can be considered truly

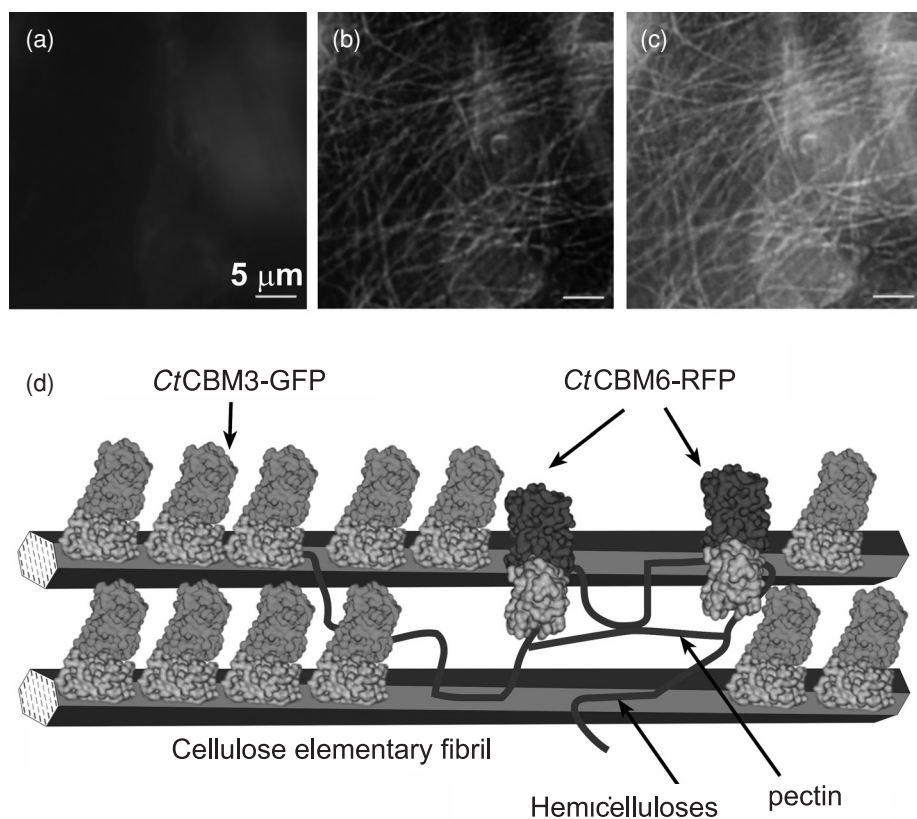


Figure 3.10 Total internal reflection fluorescence micrograph of fresh maize parenchyma cell wall labeled by CtCBM3-GFP and CtCBM6-RFP. (Modified from Ding *et al.*, 2006.) (Reproduced in color as Plate 3.)

crystalline. This hypothesis is yet to be tested due to the lack of available experimental techniques. Combined *ab initio* and force-field molecular dynamics simulations provide a unique tool to investigate the plant cell wall structure at the molecular and electronic levels. Molecular modeling will help shed light on the true nature of biomass recalcitrance at the atomic and molecular levels, thus pointing out ways to overcome this resistance to its deconstruction. Even less is understood about the structures of hemicelluloses, and the mechanisms and energetics between hemicelluloses and cellulose interactions. Hemicelluloses are polysaccharides consisting of mostly xylose and other minor sugars. Xylan and xyloglucan (XG) are the main hemicelluloses in plant cell walls. Xylan displays large structural variation and complexity. The backbone of xylan is a linear polymer of xylose linked via the β -(1-4) glycosidic bond. In higher plants, the linear backbone is substituted by a variety of side chains mainly α -L-arabinofuranosyl and α -D-glucopyranosyl uronic acid units (73). XG has a glucan backbone wherein up to 75% of the glucose (G) units are substituted at O6 with α -D-xylose (X) (74). Some of the xylose residues are then substituted at O2 with β -D-galactose (L), which can be further substituted at O2 with α -L-fucose (F). Previous molecular dynamics studies (75–79) of plant cell wall structure focused mainly on

crystalline cellulose structures with unrealistic sizes and dimensions. Furthermore, the majority of these simulations relied only on force-field-based molecular dynamics simulations (73, 74, 76–78). It is known that the accuracy of the simulations of carbohydrates will depend on the quality of the force field used. Matthews and coworkers (75) used second-generation CHARMM force fields; however, their cellulose crystal size was significantly larger than the actual size in plant cell walls. It is known that surfaces and interfaces will dominate the properties of nanostructured materials. It is thus possible that this diminished cellulose surface to volume ratio in Matthews' initial work led to some unrealistic conformational and hydrogen bonding structures in cellulose. Because the properties of cellulose and hemicelluloses are dominated by hydrogen bonding interactions, the force field method is not ideal for describing this type of interaction. In contrast, application of *ab initio*-based molecular dynamics simulations to investigate the atomic and electronic structures of crystalline cellulose I β yielded very good agreement with X-ray and neutron diffraction results, particularly for the hydrogen bonding network (79). *Ab initio* molecular dynamics with CPMD (80, 81) is another promising method to investigate the structures of celluloses and hemicelluloses in regard to their interaction mechanisms and general energetics. CPMD is capable of simulating thousands of atoms with currently available computing power (82). Combined with classical MD, it is possible to investigate the structures of the plant cell wall with up to tens of thousands of atoms.

3.5 Summary

The natural structure of modern plants (83) believed to contribute to the recalcitrance of biomass to chemical or enzymatic degradation include (i) the epidermal tissue of the plant body, particularly the cuticle and epicuticular waxes, (ii) the arrangement and density of the vascular bundles, (iii) the relative amount of sclerenchymatous (thick wall) tissue, (iv) the degree of lignification (14), (v) the warty layer covering of secondary cell walls, (vi) the structural heterogeneity and complexity of cell wall constituents, such as microfibrils and matrix polymers (84), (vii) the challenges for enzymes acting on an insoluble substrate (85), and (viii) the presence of inhibitors in cell walls or generated during the conversion processes to subsequent fermentations (86). These chemical and structural features affect liquid penetration and/or enzyme accessibility and activity, and thus the overall biomass conversion costs.

Current biomass conversion technology uses chemical pretreatments to remove hemicelluloses from the microfibrils, which in turn exposes the crystalline cellulose core rendering it more amenable to the action of cellulase enzymes. In addition, pretreatment typically breaks down the macroscopic rigidity of biomass and decreases the physical barriers to mass transport. A suite of enzymes such as cellulase, hemicellulases, and accessory enzymes are introduced to depolymerize cellulose (saccharification). During this process, the cell walls are decomposed at the molecular level, hemicelluloses and lignins are either hydrolyzed in place or allowed to migrate, and the crystalline cellulose structures are exposed and modified. Eventually, the polysaccharides are depolymerized to monomer sugars for fermentation. The technical barriers that contribute to the high cost of current biomass conversion processes have been identified as low sugar yields and low efficiency of enzyme performance. To overcome these problems, improvements of these processes rely on

further understanding of cell wall ultrastructure and the molecular mechanisms of enzyme hydrolysis.

The lignocellulose biorefinery is envisioned to comprise four major processing steps: 1) feedstock harvest and storage, 2) thermochemical pretreatment, 3) enzymatic hydrolysis, and 4) sugar fermentation to ethanol or bio-based products. Among these processes, chemical and enzymatic treatments of biomass contribute to the majority of the processing cost. An acidic chemical pretreatment step is usually conducted to depolymerize and solubilize hemicelluloses (approximately 20–40% wt/wt of biomass). This step converts hemicelluloses to monosaccharides and oligosaccharides, which can be further hydrolyzed in the later processes. Thermochemical pretreatment of biomass has long been recognized as a critical step to produce celluloses with acceptable enzymatic digestibility. Various technologies have been developed to accomplish this goal. For example, dilute sulfuric acid pretreatment at 140–200°C renders the cellulose in cell walls more accessible to enzymes (83). For dilute acid treatments (pH \sim 1.5), release of mono and oligomeric sugars from hemicellulose exhibits multi-modal kinetics. It is this slow monosaccharide release phase of chemical hemicellulose hydrolysis that directly relates to the high process conversion cost (87, 88). A number of researchers (88–93) have noted that the depolymerization of hemicellulose appears to be best described as a pair of parallel first-order reactions where one takes place at a fast rate and the other at a much slower rate. Pretreatment schemes based on alkaline explosive decomposition or organic solvent extractions have also been used with considerable success (86). The alkaline process, known as ammonia fiber expansion (AFEX), leaves the hemicellulose in place, yet renders the remaining cell walls considerably more amenable to enzymatic hydrolysis (94). At moderate pretreatment severities (95), the hemicelluloses are hydrolyzed and solubilized as monomers and oligomers; however, the yields of solubilized sugars are often unpredictable, and less than ideal (96). The improvements of chemical pretreatments now focus on increasing sugar yields and reducing the severity.

The factors that govern the pretreatment process at the level of the cell wall are not clear today. However, this process undoubtedly depends on a number of factors, such as hemicellulose composition, biomass density, the presence of non-sugar components (i.e., lignin, ash, acetyl, and uronic acids), and most importantly plant cell wall structure (i.e., types of cells, ratios of primary and secondary cell walls, as well as the macromolecular structure and arrangement of cell wall polymers).

Although it is not fully known how many enzymes are involved in cell wall deconstruction in nature, over a hundred families of glycoside hydrolases (GH) have been identified in the CAZY database (<http://afmb.cnrs-mrs.fr/CAZY/fam/acc-GH.html>). Three general categories of enzymes are considered necessary to hydrolyze native cell wall materials: cellulases, hemicellulases, and the accessory enzymes, which include hemicellulose debranching, phenolic acid esterase, and possibly lignin degrading/modifying enzymes (97). Once the hemicellulose barrier associated with cell wall microfibrils has been compromised by chemical pretreatments, cellulase enzymes can be used to hydrolyze the crystalline cellulose. Crystalline cellulose is hydrolyzed by the synergistic action of endo-acting enzymes known as endoglucanases, and exo-acting enzymes, known as exoglucanases. The endoglucanases locate surface sites along the glucan chain and insert a water molecule in the β -(1,4) bond, creating a new reducing and non-reducing chain end pair. The release of cellobiose from the cellulose is thought to occur at these new chain ends and this process considered to be the rate limiting step in cellulase action, is accomplished by exoglucanases also known as

the “processive” cellulases. At this time, studies of the synergistic reaction of cellulases are primarily based on assays on purified cellulose substrate such as Sigmacell, Avicel, or bacterial cellulose, not cell walls (98). There is no doubt that the deconstruction of the complex structures found in cell walls require a wider range of enzymes than just the cellulases; in fact, the synergistic action of many GH family enzymes as well as whole microbial cells are likely critical, and yet poorly understood (99).

Acknowledgment

This work was supported by the US Department of Energy, Office of the Biomass Program.

References

1. Cosgrove, D.J. (2001) Wall structure and wall loosening. A look backwards and forwards. *Plant Physiology*, **125**, 131–134.
2. Vincent, J.F.V. (1999) From cellulose to cell. *Journal of Experimental Biology*, **202**, 3263–3268.
3. Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., Osborne, E., Paredes, A., Persson, S., Raab, T., Vorwerk, S. & Youngs, H. (2004) Toward a systems approach to understanding plant-cell walls. *Science*, **306**, 2206–2211.
4. Doblin, M.S., Kurek, I., Jacob-Wilk, D. & Delmer, D.P. (2002) Cellulose biosynthesis in plants: From genes to rosettes. *Plant and Cell Physiology*, **43**, 1407–1420.
5. Mueller, S.C., Brown, R.M. & Scott, T.K. (1976) Cellulosic microfibrils – nascent stages of synthesis in a higher plant-cell. *Science*, **194**, 949–951.
6. Emons, A.M.C. (1991) *Role of Particle Rosettes and Terminal Globules in Cellulose Synthesis*. Marcel Dekker, New York.
7. Brown, R.M. & Saxena, I.M. (2000) Cellulose biosynthesis: A model for understanding the assembly of biopolymers. *Plant Physiology and Biochemistry*, **38**, 57–67.
8. Kimura, S., Laosinchai, W., Itoh, T., Cui, X.J., Linder, C.R. & Brown, R.M. (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. *Plant Cell*, **11**, 2075–2085.
9. Ding, S.-Y. & Himmel, M.E. (2006) The maize primary cell wall microfibril: A new model derived from direct visualization. *Journal of Agricultural and Food Chemistry*, **54**, 597–606.
10. Perrin, R.M. (2001) Cellulose: How many cellulose synthases to make a plant? *Current Biology*, **11**, R213–R216.
11. Urbanowicz, B.R., Rayon, C. & Carpita, N.C. (2004) Topology of the maize mixed linkage (1 → 3),(1 → 4)-beta-D-glucan synthase at the Golgi membrane. *Plant Physiology*, **134**, 758–768.
12. Engels, F.M. & Brice, R.E. (1985) A barrier covering lignified cell-walls of barley straw that restricts access by rumen microorganisms. *Current Microbiology*, **12**, 217–223.
13. Ha, M.A., Apperley, D.C., Evans, B.W., Huxham, M., Jardine, W.G., Vietor, R.J., Reis, D., Vian, B. & Jarvis, M.C. (1998) Fine structure in cellulose microfibrils: NMR evidence from onion and quince. *Plant Journal*, **16**, 183–190.
14. Cosgrove, D.J. (2005) Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology*, **6**, 850–861.
15. Nishiyama, Y., Sugiyama, J., Chanzy, H. & Langan, P. (2003) Crystal structure and hydrogen bonding system in cellulose I(alpha), from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, **125**, 14300–14306.

16. Nishiyama, Y., Langan, P. & Chanzy, H. (2002) Crystal structure and hydrogen-bonding system in cellulose I beta from synchrotron X-ray and neutron fiber diffraction. *Journal of the American Chemical Society*, **124**, 9074–9082.
17. Jarvis, M. (2003) Chemistry – cellulose stacks up. *Nature*, **426**, 611–612.
18. Atalla, R.H. & VanderHart, D.L. (1984) Native cellulose – a composite of 2 distinct crystalline forms. *Science*, **223**, 283–285.
19. Sugiyama, J., Persson, J. & Chanzy, H. (1991) Combined IR and electron diffraction study of the polymorphism of native cellulose. *Macromolecules*, **24**, 2461–2466.
20. Šturcová, A., His, I., Apperley, D.C., Sugiyama, J. & Jarvis, M.C. (2004) Structural details of crystalline cellulose from higher plants. *Biomacromolecules*, **5**, 1333–1339.
21. Atalla, R.H. & VanderHart, D.L. (1999) The role of solid state C-13 NMR spectroscopy in studies of the nature of native celluloses. *Solid State Nuclear Magnetic Resonance*, **15**, 1–19.
22. Ranby, B.G. (1951) The colloidal properties of cellulose micelles. *Discussions Faraday Society*, **11**, 158–164.
23. Åkerholm, M., Hinterstoisser, B. & Salména, L. (2004) Characterization of the crystalline structure of cellulose using static and dynamic FT-IR spectroscopy. *Carbohydrate Research*, **339**, 569–578.
24. Vietor, R.J., Newman, R.H., Ha, M.A., Apperley, D.C. & Jarvis, M.C. (2002) Conformational features of crystal-surface cellulose from higher plants. *Plant Journal*, **30**, 721–731.
25. Bootten, T.J., Harris, P.J., Melton, L.D. & Newman, R.H. (2004) Solid-state C-13-NMR spectroscopy shows that the xyloglucans in the primary cell walls of mung bean (*Vigna radiata* L.) occur in different domains: A new model for xyloglucan–cellulose interactions in the cell wall. *Journal of Experimental Botany*, **55**, 571–583.
26. Arioli, T., Peng, L.C., Betzner, A.S., Burn, J., Wittke, W., Herth, W., Camilleri, C., Hofte, H., Plazinski, J., Birch, R., Cork, A., Glover, J., Redmond, J. & Williamson, R.E. (1998) Molecular analysis of cellulose biosynthesis in Arabidopsis. *Science*, **279**, 717–720.
27. Vance, C.P., Kirk, T.K. & Sherwood, R.T. (1980) Lignification as a mechanism of disease resistance. *Annual Review of Phytopathology*, **18**, 259–288.
28. Whetten, R.W., MacKay, J.J. & Sederoff, R.R. (1998) Recent advances in understanding lignin biosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, **49**, 585–609.
29. Boerjan, W., Ralph, J. & Baucher, M. (2003) Lignin biosynthesis. *Annual Review of Plant Biology*, **54**, 519–546.
30. Baucher, M., Monties, B., Van Montagu, M. & Boerjan, W. (1998) Biosynthesis and genetic engineering of lignin. *Critical Reviews in Plant Sciences*, **17**, 125–197.
31. Donaldson, L.A. (2001) Lignification and lignin topochemistry – an ultrastructural view. *Phytochemistry*, **57**, 859–873.
32. Ohad, I., Danon, D. & Hestrin, S. (1962) Synthesis of cellulose by *Acetobacter xylium*: V. Ultrastructure of polymer. *Journal of Cell Biology*, **12**, 31–46.
33. Ohad, I. & Danon, D. (1964) On the dimensions of cellulose microfibrils. *Journal of Cell Biology*, **22**, 302–305.
34. McCann, M.C., Wells, B. & Roberts, K. (1990) Direct visualization of cross-links in the primary plant-cell wall. *Journal of Cell Science*, **96**, 323–334.
35. Marga, F., Grandbois, M., Cosgrove, D.J. & Baskin, T.I. (2005) Cell wall extension results in the coordinate separation of parallel microfibrils: Evidence from scanning electron microscopy and atomic force microscopy. *Plant Journal*, **43**, 181–190.
36. Carpita, N.C., Defernez, M., Findlay, K., Wells, B., Shoue, D.A., Catchpole, G., Wilson, R.H. & McCann, M.C. (2001) Cell wall architecture of the elongating maize coleoptile. *Plant Physiology*, **127**, 551–565.
37. Morris, V.J., Kirby, A.R. & Gunning, A.P. (1999) *Atomic Force Microscopy for Biologists*. Imperial College Press, London.

38. van der Wel, N.N., Putman, C.A.J., vanNoort, S.J.T., deGroot, B.G. & Emons, A.M.C. (1996) Atomic force microscopy of pollen grains, cellulose microfibrils, and protoplasts. *Protoplasma*, **194**, 29–39.
39. Davies, L.M. & Harris, P.J. (2003) Atomic force microscopy of microfibrils in primary cell walls. *Planta*, **217**, 283–289.
40. Kirby, A.R., Gunning, A.P., Waldron, K.W., Morris, V.J. & Ng, A. (1996) Visualization of plant cell walls by atomic force microscopy. *Biophysical Journal*, **70**, 1138–1143.
41. Engel, A., Lyubchenko, Y. & Muller, D. (1999) Atomic force microscopy: A powerful tool to observe biomolecules at work. *Trends in Cell Biology*, **9**, 77–80.
42. Thimm, J.C., Burritt, D.J., Ducker, W.A. & Melton, L.D. (2000) Celery (*Apium graveolens* L.) parenchyma cell walls examined by atomic force microscopy: Effect of dehydration on cellulose microfibrils. *Planta*, **212**, 25–32.
43. Morris, V.J., Gunning, A.P., Kirby, A.R., Round, A., Waldron, K. & Ng, A. (1997) Atomic force microscopy of plant cell walls, plant cell wall polysaccharides and gels. *International Journal of Biological Macromolecules*, **21**, 61–66.
44. Feijo, J. & Moreno, N. (2004) Imaging plant cells by two-photon excitation. *Protoplasma*, **223**, 1–22.
45. Squier, J.A., Muller, M., Brakenhoff, G.J. & Wilson, K.R. (1998) Third harmonic generation microscopy. *Optics Express*, **3**, 315–324.
46. Book, L.D., Cheng, J.X., Volkmer, A. & Xie, X.S. (2001) CARS microscopy of living cells with high spectral resolution and sensitivity. *Biophysical Journal*, **80**, 171a.
47. Legare, F., Evans, C.L., Ganikhanov, F. & Xie, X.S. (2006) Towards CARS endoscopy. *Optics Express*, **14**, 4427–4432.
48. Ishii, Y., Kimura, Y., Kitamura, K., Tanaka, H., Wazawa, T. & Yanagida, T. (2000) Imaging and nano-manipulation of single actomyosin motors at work. *Clinical and Experimental Pharmacology and Physiology*, **27**, 229–237.
49. Sako, Y. & Uyemura, T. (2002) Total internal reflection fluorescence microscopy for single-molecule imaging in living cells. *Cell Structure and Function*, **27**, 357–365.
50. Lukyanov, K.A., Chudakov, D.M., Lukyanov, S. & Verkhusha, V.V. (2005) Innovation: Photoactivatable fluorescent proteins. *Nature Review Molecular Cell Biology*, **6**, 885–891.
51. Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J. & Hess, H.F. (2006) Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, **313**, 1642–1645.
52. Rust, M.J., Bates, M. & Zhuang, X. (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, **3**, 793–795.
53. Lippincott-Schwartz, J. & Patterson, G.H. (2003) Development and use of fluorescent protein markers in living cells. *Science*, **300**, 87–91.
54. Patterson, G.H. & Lippincott-Schwartz, J. (2002) A photoactivatable GFP for selective photolabeling of proteins and cells. *Science*, **297**, 1873–1877.
55. Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H. & Miyawaki, A. (2002) An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 12651–12656.
56. Karasawa, S., Araki, T., Yamamoto-Hino, M. & Miyawaki, A. (2003) A green-emitting fluorescent protein from Galaxiidae coral and its monomeric version for use in fluorescent labeling. *Journal of Biological Chemistry*, **278**, 34167–34171.
57. Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Rocker, C., Salih, A., Spindler, K.D. & Nienhaus, G.U. (2004) EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15905–15910.

58. Ando, R., Mizuno, H. & Miyawaki, A. (2004) Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science*, **306**, 1370–1373.
59. Ding, S.Y., Xu, Q., Ali, M.K., Baker, J.O., Bayer, E.A., Barak, Y., Lamed, R., Sugiyama, J., Rumbles, G. & Himmel, M.E. (2006) Versatile derivatives of carbohydrate-binding modules for imaging of complex carbohydrates approaching the molecular level of resolution. *Biotechniques*, **41**, 435–443.
60. McCartney, L., Marcus, S.E. & Knox, J.P. (2005) Monoclonal antibodies to plant cell wall xylans and arabinoxylans. *Journal of Histochemistry and Cytochemistry*, **53**, 543–546.
61. Hahn, M.G., Freshour, G., Chafin, D. & Braga, M.R. (2005) Monoclonal antibody tools for studies of plant cell wall structure and dynamics. *Abstracts of Papers of the American Chemical Society*, **229**, U295.
62. Trethewey, J.A.K., Campbell, L.M. & Harris, P.J. (2005) (1 → 3), (1 → 4)-beta-D-Glucans in the cell walls of the poales (sensu lato): An immunogold labeling study using a monoclonal antibody. *American Journal of Botany*, **92**, 1660–1674.
63. Clausen, M.H., Ralet, M.C., Willats, W.G.T., McCartney, L., Marcus, S.E., Thibault, J.F. & Knox, J.P. (2004) A monoclonal antibody to feruloylated-(1 → 4)-beta-D-galactan. *Planta*, **219**, 1036–1041.
64. Willats, W.G.T., Steele-King, C.G., McCartney, L., Orfila, C., Marcus, S.E. & Knox, J.P. (2000) Making and using antibody probes to study plant cell walls. *Plant Physiology and Biochemistry*, **38**, 27–36.
65. Takeuchi, M., Takabe, K. & Fujita, M. (2005) Immunolocalization of an anionic peroxidase in differentiating poplar xylem. *Journal of Wood Science*, **51**, 317–322.
66. Joseleau, J.P., Faix, O., Kuroda, K.I. & Ruel, K. (2004) A polyclonal antibody directed against syringylpropane epitopes of native lignins. *Comptes Rendus Biologies*, **327**, 809–815.
67. Musel, G., Schindler, T., Bergfeld, R., Ruel, K., Jacquet, G., Lapiere, C., Speth, V. & Schopfer, P. (1997) Structure and distribution of lignin in primary and secondary cell walls of maize coleoptiles analyzed by chemical and immunological probes. *Planta*, **201**, 146–159.
68. Gunnarsson, L.C., Zhou, Q., Montanier, C., Karlsson, E.N., Brumer, H. & Ohlin, M. (2006) Engineered xyloglucan specificity in a carbohydrate-binding module. *Glycobiology*, **16**, 1171–1180.
69. McCartney, L., Gilbert, H.J., Bolam, D.N., Boraston, A.B. & Knox, J.P. (2004) Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. *Analytical Biochemistry*, **326**, 49–54.
70. Bayer, E.A., Belaich, J.P., Shoham, Y. & Lamed, R. (2004) The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology*, **58**, 521–554.
71. Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. (2004) Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochemical Journal*, **382**, 769–781.
72. Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M. & Teeri, T.T. (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 484–489.
73. Mazeau, K., Moine, C., Krausz, P. & Gloaguen, V. (2005) Conformational analysis of xylan chains. *Carbohydrate Research*, **340**, 2752–2760.
74. Hanus, J. & Mazeau, K. (2006) The xyloglucan–cellulose assembly at the atomic scale. *Biopolymers*, **82**, 59–73.
75. Matthews, J.F., Skopec, C.E., Mason, P.E., Zuccato, P., Torget, R.W., Sugiyama, J., Himmel, M.E. & Brady, J.W. (2006) Computer simulation studies of microcrystalline cellulose I beta. *Carbohydrate Research*, **341**, 138–152.
76. Vietor, R.J., Mazeau, K., Lakin, M. & Perez, S. (2000) A priori crystal structure prediction of native celluloses. *Biopolymers*, **54**, 342–354.
77. Mazeau, K. & Heux, L. (2003) Molecular dynamics simulations of bulk native crystalline and amorphous structures of cellulose. *Journal of Physical Chemistry B*, **107**, 2394–2403.

78. Mazeau, K. (2005) Structural micro-heterogeneities of crystalline I beta-cellulose. *Cellulose*, **12**, 339–349.
79. Qian, X.H., Ding, S.Y., Nimlos, M.R., Johnson, D.K. & Himmel, M.E. (2005) Atomic and electronic structures of molecular crystalline cellulose I beta: A first-principles investigation. *Macromolecules*, **38**, 10580–10589.
80. Car, R. & Parrinello, M. (1985) Unified approach for molecular-dynamics and density-functional theory. *Physical Review Letters*, **55**, 2471–2474.
81. Car-Parrinello Molecular Dynamics. Available from: <http://www.cpmc.org>.
82. Andreoni, W. & Curioni, A. (2000) New advances in chemistry and materials science with CPMD and parallel computing. *Parallel Computing*, **26**, 819–842.
83. Himmel, M.E., Ding, S.-Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W. & Foust, T.D. (2007) Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*, **315**, 804–807.
84. Iiyama, K., Lam, T.B.T. & Stone, B.A. (1994) Covalent cross-links in the cell-wall. *Plant Physiology*, **104**, 315–320.
85. Himmel, M.E., Ruth, M.F. & Wyman, C.E. (1999) Cellulase for commodity products from cellulosic biomass. *Current Opinion in Biotechnology*, **10**, 358–364.
86. Wyman, C.E., Dale, B.E., Elander, R.T., Holtzapple, M., Ladisch, M.R. & Lee, Y.Y. (2005) Coordinated development of leading biomass pretreatment technologies. *Bioresource Technology*, **96**, 1959–1966.
87. Liu, C.G. & Wyman, C.E. (2004) The effect of flow rate of very dilute sulfuric acid on xylan, lignin, and total mass removal from corn stover. *Industrial and Engineering Chemistry Research*, **43**, 2781–2788.
88. Esteghlalian, A., Hashimoto, A.G., Fenske, J.J. & Penner, M.H. (1997) Modeling and optimization of the dilute-sulfuric-acid pretreatment of corn stover, poplar and switchgrass. *Bioresource Technology*, **59**, 129–136.
89. Shiang, M., Linden, J.C., Mohagheghi, A., Grohmann, K. & Himmel, M.E. (1991) Regulation of cellulase synthesis in acidothermus-cellulolyticus. *Biotechnology Progress*, **7**, 315–322.
90. Kim, S.B. & Lee, Y.Y. (1987) Kinetics in acid-catalyzed hydrolysis of hardwood hemicellulose. *Biotechnology and Bioengineering Symposium*, **17**, 71–84.
91. Maloney, M.T., Chapman, T.W. & Baker, A.J. (1985) Dilute acid-hydrolysis of paper birch – kinetics studies of xylan and acetyl-group hydrolysis. *Biotechnology and Bioengineering*, **27**, 355–361.
92. Mayans, O., Scott, M., Connerton, I., Gravesen, T., Benen, J., Visser, J., Pickersgill, R. & Jenkins, J. (1997) Two crystal structures of pectin lyase A from aspergillus reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases. *Structure*, **5**, 677–689.
93. Chen, R., Lee, Y.Y. & Torget, R. (1996) Kinetic and modeling investigation on two-stage reverse-flow reactor as applied to dilute-acid pretreatment of agricultural residues. *Applied Biochemistry and Biotechnology*, **57/58**, 133–146.
94. Teymouri, F., Laureano-Perez, L., Alizadeh, H. & Dale, B.E. (2005) Optimization of the ammonia fiber explosion (AFEX) treatment parameters for enzymatic hydrolysis of corn stover. *Bioresource Technology*, **96**, 2014–2018.
95. Holland, N., Holland, D., Helentjaris, T., Dhugga, K.S., Xoconostle-Cazares, B. & Delmer, D.P. (2000) A comparative analysis of the plant cellulose synthase (CesA) gene family. *Plant Physiology*, **123**, 1313–1323.
96. Torget, R., Himmel, M.E. & Grohmann, K. (1991) Dilute sulfuric-acid pretreatment of hardwood bark. *Bioresource Technology*, **35**, 239–246.
97. Himmel, M.E., Adney, W.S., Baker, J.O., Elander, R., McMillan, J.D., Nieves, R.A., Sheehan, J.J., Thomas, S.R., Vinzant, T.B. & Zhang, M. (1997) Advanced bioethanol production technologies: A perspective. *Fuels and Chemicals from Biomass*, **666**, 2–45.

98. Baker, J.O., Thomas, S.R., Adney, W.S., Nieves, R.A. & Himmel, M.E. (1994) The cellulase synergistic effect – binary and ternary-systems. *Abstracts of Papers of the American Chemical Society*, **207**, 50.
99. Lu, Y.P., Zhang, Y.H.P. & Lynd, L.R. (2006) Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 16165–16169.

Chapter 4

Chemistry and Molecular Organization of Plant Cell Walls

Philip J. Harris and Bruce A. Stone

Abbreviations used: Apif, Apiofuranosyl; Araf, Arabinofuranosyl; AGP, Arabinogalactan-protein; Arap, Arabinopyranosyl; pCA, *p*-Coumaric acid; DDFA, Dehydrodiferulic acid; DDQ, 2,3-dichloro-5,6-dicyano-1,4-benzoquinone; FA, Ferulic acid; Fucp, Fucopyranosyl; Galp, Galactopyranosyl; Glcp, Glucopyranosyl; GlcpA, Glucopyranuronosyl; GalpA, Galactopyranuronosyl; GAX, Glucuronoarabinoxylan; GRP, Glycine-rich protein; HCA, Hydroxycinnamic acid; HGA, Homogalacturonan; HRGP, Hydroxyproline-rich glycoprotein; 4-*O*-methyl- α -D-GlcpA, 4-*O*-methyl- α -D-glucopyranosyluronic acid; Rhap, Rhamnopyranosyl; RG-I, Rhamnogalacturonan I; RG-II, Rhamnogalacturonan II; Rhap, Rhamnopyranosyl; SA, Sinapic acid; XGA, Xylogalacturonan; XG, Xyloglucan; Xylp, Xylopyranosyl.

4.1 Introduction

The cell walls of seed plants (angiosperms and gymnosperms) represent an enormous store of fermentable carbohydrate that is a potential source of ethanol. However, this carbohydrate is in the form of high molecular weight cellulose and the accompanying non-cellulosic polysaccharides, which are intimately associated both non-covalently and covalently with one another and often with non-carbohydrate polymers particularly lignins, and other polymers such as proteins, and in some situations suberin and cutin. These associations must be broken to allow access of polysaccharide degrading enzymes to their substrates during the digestion phase of bioethanol production in which fermentable sugars (monosaccharides) are released from the plant feedstocks.

Two types of cell walls are recognized, primary and secondary (1–3). Primary walls are formed while cells are still developing and enlarging and for many cell types, e.g., some parenchyma cells, the primary wall is the only wall. Primary walls are typically non-lignified and their thickness in mature cells depends on the cell type. Secondary walls, e.g., in sclerenchyma fibers, are deposited on the primary wall after the cells are fully expanded and at maturity the entire wall (primary and secondary) is typically lignified. The middle lamella, the interfacial layer between adjacent cells, which develops from the cell plate present at division, is also typically lignified. Although less common, cell types occur that have both primary and secondary walls but that are not lignified, or only slightly lignified, for example,

the fibers of tension wood (in hardwoods), bast (phloem) fibers from the stems of flax (*Linum usitatissimum*), hemp (*Cannabis sativa*), ramie (*Boehmeria nivea*) and nettle (*Urtica dioica*), and the hairs of cotton (*Gossypium* spp.) seeds (4). Many of the parenchyma cells in stems of grasses have secondary walls and the entire wall is lignified (5).

This chapter provides a background to the discussion on current methods for disassociating polysaccharides from plant cell walls to make them accessible to depolymerizing enzymes (see Chapter 14) and on the specificities and action patterns of the polysaccharide depolymerizing enzymes that may be used to release monosaccharides for fermentation (see Chapter 10). In particular, we review the chemistry of cell wall polysaccharides, their associations with one another and with non-carbohydrate polymers, and models of cell wall architecture. We shall focus particularly on primary and lignified secondary cell walls of vegetative organs of common groups of plants that have been suggested as major biomass sources for bioethanol production. Of particular importance are the herbaceous angiosperms, especially the grasses, including the cereals, which form the monocotyledon family Poaceae. This family yields grain-milling residues, e.g., wheat (*Triticum aestivum*) bran, and crop residues, e.g., wheat straw and maize (*Zea mays*) stover. In addition, perennial grasses such as *Miscanthus* (*Miscanthus* spp.), switchgrass (*Panicum virgatum*), giant reed (*Arundo donax*), and reed canarygrass (*Phalaris arundinacea*) have been proposed in the United States (6) or selected by the European Union (7) as being particularly promising as energy crops. The grass family forms part of a major group of monocotyledons known as commelinid monocotyledons that are characterized by the presence of ester-linked ferulic acid in their primary walls. We refer to other monocotyledons as non-commelinid monocotyledons (8).

In the rest of the angiosperms, the term eudicotyledon (or true dicotyledons) is now used to refer to most of the group previously known as the dicotyledons. Some herbaceous eudicotyledons, such as the forage legume alfalfa (*Medicago sativa*), have also been considered as energy crops (9). Hardwoods (woody angiosperms), e.g., poplars (*Populus* spp.) and willows (*Salix* spp.), and softwoods (coniferous gymnosperms), e.g., spruce (*Picea abies*), are also potential feedstocks.

4.2 Chemistry of cell wall polymers

4.2.1 Chemistry of cell wall polysaccharides

Cell wall polysaccharides are of two types: the first is the stereo-regular homopolymer type among which cellulose that makes up the microfibrillar phase of all cell walls is the premier example, and the second comprises the stereo-irregular, non-cellulosic polysaccharides represented by the (1→3,1→4)-β-D-glucans, heteroxylans, heteroglucans, heteromannans, and the pectins (pectic polysaccharides) which are found in the matrix phase of the wall. The chemistry and occurrence of these two polysaccharide types is described in the following sections. Table 4.1 summarizes the cell wall components of the vegetative parts of target “cellulosic” ethanol feedstocks. It should be noted that the stereo-irregular polysaccharides other than the pectins are loosely referred to as hemicelluloses in many publications. We prefer to use the term non-cellulosic polysaccharides to describe collectively all these molecules and to refer to the individual polysaccharides specifically by their descriptive chemical names (8). The extension of the term to “hemicellulase” further obfuscates the

Table 4.1 Compositions of the cell walls of vegetative parts of “cellulosic” ethanol feedstocks

Plant group	Wall type	
	Primary walls	Secondary walls
Angiosperms		
Eudicotyledons		
e.g., Hardwoods [e.g., willows (<i>Salix</i> spp.) <i>Eucalyptus</i> spp.] and legumes	Fibrillar phase: cellulose Matrix phase: Pectic polysaccharides > xyloglucans > heteroxylans and (galacto)glucomannans (minor) Ester-linked ferulic and small proportions of <i>p</i> -coumaric acid in “core” Caryophyllales. Known to be linked to RG-I side chains in Amaranthaceae	(Hardwoods) Fibrillar phase: cellulose Matrix phase: 4- <i>O</i> -methyl-glucurono-xylans > glucomannans Lignins: Syringyl and guaiacyl; small proportions of <i>p</i> -hydroxyphenyl units. Acylated with <i>p</i> -hydroxybenzoic acid in Salicaceae and acetic acid in <i>Hibiscus cannabinus</i>
Monocotyledons		
Non-commelinids		
e.g., Sisal (<i>Agave sisalana</i>)	Fibrillar phase: cellulose Matrix phase: As eudicotyledons	Fibrillar phase: cellulose Matrix phase: As eudicotyledons
Commelinids		
e.g., Grasses and cereals (Poaceae), palms (Arecales)	Fibrillar phase: cellulose ^a Matrix phase: Glucuronoarabinoxylans (GAXs) > pectic polysaccharides and xyloglucans [(1→3,1→4)-β-glucans in some Poales families, including Poaceae] Ferulic, small proportions of <i>p</i> -coumaric, and sometimes sinapic acids, ester-linked to GAXs	Fibrillar phase: cellulose Matrix phase: Glucuronoarabinoxylans > (galacto-)glucomannans (minor) Lignins: As in eudicotyledons. Acylated with <i>p</i> -coumaric acid in Poaceae and Musaceae, and with <i>p</i> -hydroxybenzoic acid in Arecales
Gymnosperms		
e.g., Coniferous gymnosperms, e.g., spruce (<i>Picea abies</i>)	Fibrillar phase: cellulose ^b Matrix phase: As eudicotyledons Ester-linked ferulic and <i>p</i> -coumaric acids in all families	(Softwoods) Fibrillar phase: cellulose Matrix phase: (Galacto-)glucomannans > 4- <i>O</i> -methyl-glucuronoarabinoxylans Lignins: Usually mostly guaiacyl units; also <i>p</i> -hydroxyphenyl units in compression wood

^a The primary walls of palms (Arecales) are an exception; their matrix polysaccharides are pectic polysaccharides > xyloglucans > glucuronoarabinoxylans (minor).

^b Detailed data on the matrix polysaccharides are available only for conifers (Coniferales) and *Ginkgo biloba* (Ginkgoales).

situation since the enzymes hydrolyzing non-cellulosic polysaccharides are specific and cannot be grouped together (see Chapter 10).

4.2.1.1 Microfibrillar phase polysaccharide: cellulose [(1→4)-β-D-glucan]

Cellulose is a ubiquitous polysaccharide component of plant cell walls. The cellulose content of cell walls of vegetative organs varies from ~20 to 40%. The lower end of the range applies to non-lignified primary walls, whereas the upper end refers to lignified secondary walls.

Cellulose molecules are long, unbranched chains of β-D-glucopyranose (β-D-Glcp) residues joined by (1→4)-glucosidic linkages (Figure 4.1). The average degree of polymerization is about 6000 glucose units in primary walls and up to 14 000 in secondary walls. The molecules have an extended, ribbon-like conformation that allows parallel packing of the chains into three-dimensional microfibrillar aggregates stabilized by extensive intermolecular hydrogen bonding and van der Waals interactions. The microfibrils have diameters of ~2–4 nm in both primary walls and lignified secondary walls of softwoods and hardwoods, but in the secondary walls of both types of woods the microfibrils apparently form aggregates ~14–23 nm in diameter (10–13). Although the microfibrils of primary walls have often been assumed to consist of 36 individual cellulose molecules, solid-state ¹³C NMR studies have indicated fewer molecules (~20–25), mostly packed into highly ordered arrays (14, 15). However, the molecular conformation and hydrogen bonding arrangement is different on the surfaces (16). (The structure of plant cell wall cellulose and the organization of the molecular chains in microfibrils are discussed in detail in Chapter 6, Structures of Plant Cell Wall Celluloses.)

4.2.1.2 Matrix phase polysaccharides

4.2.1.2.1 (1→3,1→4)-β-D-GLUCANS

(1→3,1→4)-β-D-Glucans in seed plants (angiosperms and gymnosperms) are found exclusively in the walls of the grasses and certain related families (3, 17). Although the largest concentrations, up to 75%, of these polysaccharides occur in walls of the starchy endosperm of cereal grains such as barley (*Hordeum vulgare*) and oats (*Avena sativa*), they also occur, in lower concentrations in the primary walls of their vegetative organs (18, 19). For example, Smith and Harris (19) found 8.8% in primary walls from perennial ryegrass (*Lolium perenne*) stems.

The (1→3,1→4)-β-D-glucans are linear, unbranched polymers in which the β-D-Glcp residues are joined by both (1→3)- and (1→4)-glucosidic linkages. Single (1→3)-linkages are separated by two or more (1→4)-linkages (Figure 4.1) and regions of two or three adjacent (1→4)-linkages predominate with longer (1→4)-β-D-glucosides (DP 5–14) accounting for <10% by weight of the molecules.

4.2.1.2.2 HETEROGLUCANS (xyloglucans)

Xyloglucans are components of the primary walls of most seed plants examined so far. Fucogalactoxyloglucans are the commonest structural type and occur in the primary walls of coniferous gymnosperms, eudicotyledons and in the non-commelinid monocotyledons, at least in onion (*Allium cepa*) and garlic (*A. sativa*) (3). These xyloglucans have a repeating backbone subunit consisting of four (1→4)-linked β-D-Glcp residues, with the three Glcp residues nearest the non-reducing end each bearing a single (1→6)-linked

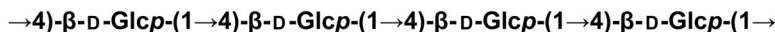
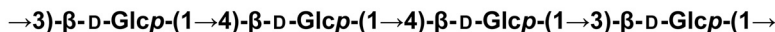
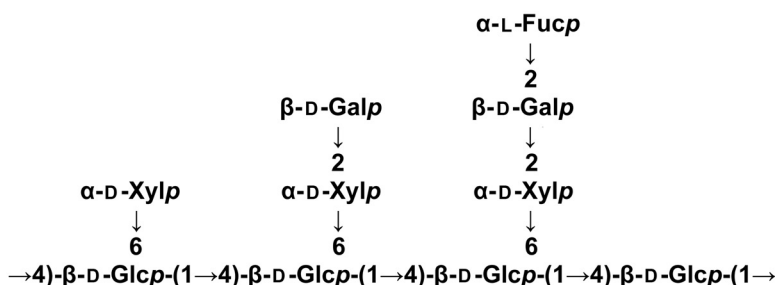
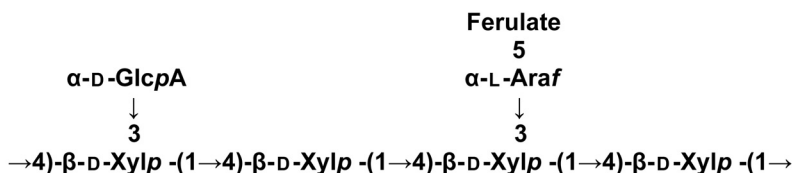
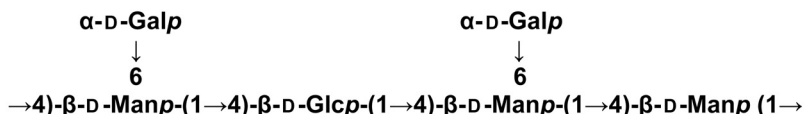
Cellulose [(1→4)-β-Glucan]**(1→3, 1→4)-β-Glucan****Heteroglucan [Fucogalactoxyloglucan]****Heteroxylan [Glucuronarabinoxylan]****Heteromannan [Galactoglucomannan]**

Figure 4.1 Structures of cellulose, (1→3,1→4)-β-D-glucan, a heteromannan (galactoglucomannan), a heteroglucan (fucogalactoxyloglucan), and a heteroxylan (glucuronarabinoxylan).

α-D-xylopyranosyl (D-Xylp) residue (Figure 4.1). About half of the subunits have a L-fucopyranosyl (L-Fucp)-D-galactopyranosyl (D-Galp) side chain attached to the Xylp residue closest to the reducing end; about one half of the subunits with this side chain also have a Galp residue attached to the adjacent Xylp residue (3). In the eudicotyledon tobacco/tomato family (Solanaceae) and related families, the xyloglucans are arabinoxyloglucans that contain L-arabinofuranosyl (L-Araf) but no L-fucopyranosyl (L-Fucp) residues (3, 20, 21). Fucose is also usually absent from the xyloglucans of the grasses; their xyloglucans also contain less xylose and much less galactose than the fucogalactoxyloglucans (3, 22, 23).

Acetyl groups also occur on xyloglucans, for example, in fucogalactoxyloglucans, the D-Galp residues may have an acetyl group attached to C(O)6 (24). In the *Nicotiana plumbaginifolia* (Solanaceae) XG, acetyl groups are found on 44% of the C(O)6 of the Glcp backbone residues not substituted with Xylp residues and at C(O)5 of 15% of the terminal Araf residues (25).

4.2.1.2.3 HETEROXYLANS (e.g., *O*-acetyl-(4-*O*-methylglucurono)xylans)

Heteroxylans are the predominant non-cellulosic polysaccharides in many types of cell walls. They all have a basic backbone chain of β -D-Xylp residues, linked through (1 \rightarrow 4)-glycosidic linkages, and substituted by various monosaccharide and oligosaccharide side chains (2, 3, 26). Heteroxylans are the major non-cellulosic polysaccharides in the lignified secondary walls of all eudicotyledons, including the hardwoods where they constitute 10–35% of the wood, and possibly the lignified secondary walls of all non-commelinid monocotyledons. In all these walls, the heteroxylans have 4-*O*-methyl- α -D-glucopyranosyluronic acid (4-*O*-methyl- α -D-GlcpA) residues attached to about every tenth Xylp residue (4-*O*-methylglucuronoxylans) (Figure 4.1) primarily to the C(O)2 and to a lesser extent the C(O)3 position of the Xylp residues. Acetyl groups are esterified to the Xylp residues; mostly at the C(O)2, but a small proportion at C(O)3, with \sim 3.5–7 acetyl groups per ten Xylp residues (27). However, variation in the structures of 4-*O*-methylglucuronoxylans occurs. For example, in the heteroxylan from the wood of the eudicotyledon Tasmanian blue gum (*Eucalyptus globulus*), some of the 4-*O*-methylglucuronic acid residues are substituted at C(O)2 with α -D-Galp (28).

In softwoods, heteroxylans are usually the second most abundant non-cellulosic polysaccharides, making up \sim 7–15% of the wood (27). As in hardwoods, these heteroxylans have 4-*O*-methyl- α -D-GlcpA residues attached predominantly at the C(O)2 position (about one residue every 5–6 Xylp residues), but they also have α -L-Araf residues attached mainly at the C(O)3 position of Xylp residues (about one residue in every 8 Xylp residues) (Figure 4.1) and are usually referred to as arabino(4-*O*-methylglucurono)xylans. Softwood heteroxylans are not acetylated.

In the vegetative organs of grasses, heteroxylans with similar structures to gymnosperm arabino(4-*O*-methylglucurono)xylans are usually the most abundant non-cellulosic polysaccharides in both the primary and lignified secondary walls. These polysaccharides have single α -L-Araf and α -D-GlcAp (or its 4-*O*-methyl derivative) residues linked at C(O)3 and C(O)2, respectively, to the xylose residues of the (1 \rightarrow 4)- β -D-Xylp backbone (Figure 4.1) and have traditionally been referred to as glucuronoarabinoxylans (GAXs) (3, 18, 29–31). Oligosaccharide side chains also occur that contain a number of monosaccharides including the following: L-Araf, D-Xylp, D-Galp, and L-Galp. Ferulic acid (FA), small amounts of *p*-coumaric (*p*CA) and sometimes sinapic acid (SA), are esterified by their carboxyl groups to the C(O)5 hydroxyl of some of the arabinosyl residues, including those occurring singly and in oligosaccharides (32, 33). In GAXs from the walls of perennial ryegrass, and barley and wheat straw, alkali labile substituents are attached at C(O)2 and C(O)3 on 50% of the Xylp residues of the main chain. Acetyl groups could account for 50–70% of the substitutions (34). In GAXs from the walls of bamboo (*Phyllostachys edulis*) shoots, acetyl groups are on the Araf residues (35). The acetyl substituents on wheat bran xylans are more labile than the hydroxycinnamic acid (HCA) esters, but no detectable bound acetate remained after extraction with 1 M KOH (36). GAXs in lignified secondary walls have backbones with a lower overall degree of substitution than GAXs in primary walls.

GAXs with similar structures to those of the grasses are also major non-cellulosic polysaccharides of the primary walls of all commelinid monocotyledons except for the palms (Arecaceae) (3, 37–40). They are also probably the major non-cellulosic polysaccharides of lignified secondary walls of all commelinid monocotyledons (2, 3). Small proportions of heteroxylans have been reported in the primary cell walls of eudicotyledons. In the walls of sycamore suspension-cultured cells, GAXs have side chains all linked at C(O)2 to the xylose residues of the (1→4)- β -D-Xylp backbone (41). In white clover (*Trifolium repens*), 75% of the Xylp main chain residues carry acetyl substituents (34).

4.2.1.2.4 HETEROMANNANS (*galactoglucomannans and glucomannans*)

Galactoglucomannans are found particularly in the lignified secondary walls of coniferous gymnosperms (softwoods), where they are usually the predominant non-cellulosic polysaccharides, comprising ~12–18% of the wood (2, 3, 42, 43). They have a linear chain of (1→4)-linked β -D-mannopyranosyl (Manp) and β -D-Glcp residues, randomly arranged, with α -D-Galp residues attached by (1→6)-linkages to both the Manp and Glcp residues (27) (Figure 4.1). The ratio of Glc:Man:Gal varies from 1:3–4:0.1–1.0. In galactoglucomannans from spruce wood, more of the D-Galp residues are attached to Manp than Glcp residues (44). In this species, O-acetyl groups are attached mostly to some of the Manp residues at the C(O)2 and C(O)3 positions (43, 45). However, in the galactoglucomannans from wood of the Parana pine (*Araucaria angustifolia*), O-acetyl groups are reported to be attached at the C(O)3 position of 15.6% of Manp and 6.4% of Glcp residues (23).

Glucomannans are usually the second most abundant non-cellulosic polysaccharides in the walls of hardwoods, making up ~2–5% of the wood (27). They have backbone structures similar to the galactoglucomannans of softwoods, but lack Galp residues and usually a ratio of Glc:Man of ~1:2 although those from birch (*Betula* spp.) have a ratio of ~1:1. O-Acetylated glucomannans have been isolated from aspen (*Populus tremula*) and birch wood and, as with the galactoglucomannans of the softwood spruce, the acetyl groups are attached to the C(O)2 and C(O)3 positions of some of the Manp residues (46).

Galactoglucomannans also occur in the primary cell walls of most seed plants (angiosperms and gymnosperms), but usually only in small proportions. Structural characterizations of galactoglucomannans obtained either from primary walls of various eudicotyledon suspension-cultured cells or from the extracellular culture medium have been made (47–50). Unlike the heteromannans from secondary walls, they contain about equal proportions of Glcp, Manp, and Galp, have alternating β -D-Glcp and β -D-Manp residues in the backbone, and either single α -D-Galp residues or β -D-Galp-(1→2)- α -D-Galp residues attached at C(O)6 of the Manp residues. A similar galactoglucomannan but with a Glc:Man:Gal ratio of 1:1:0.5 has been characterized from primary walls of kiwifruit (*Actinidia deliciosa*) outer pericarp (51).

4.2.1.2.5 PECTIC POLYSACCHARIDES (*pectins*)

The pectic polysaccharides (pectins) are a family of complex polysaccharides that occur in the primary cell walls of all seed plants, where they are major components except in the walls of the grasses and other commelinid monocotyledons other than the palms (Arecaceae). The polysaccharides have a block or domain structure with four commonly occurring domains: homogalacturonan (HG), rhamnogalacturonan I (RG-I), rhamnogalacturonan II (RG-II), and xylogalacturonan (XGA) (52–54). Homogalacturonan (Figure 4.2), which is usually the most abundant domain, is composed of linear chains of α -D-galacturonic acid (α -D-GalAp)

$$\rightarrow 4)-\alpha\text{-D-GalpA}-(1\rightarrow 4)-\alpha\text{-D-GalpA}-(1\rightarrow 4)-\alpha\text{-D-GalpA}-(1\rightarrow 4)-\alpha\text{-D-GalpA}-(1\rightarrow$$

→2)-α-L-Rhap-(1→4)-α-D-GalpA-(1→2)-α-L-Rhap-(1→4)-α-D-GalpA-(1→
4
↑
(1→5)-α-arabinan, (1→4)-β-galactan,
arabino-4-galactan or arabino-3,6-galactan

$$\begin{array}{c}
 \alpha\text{-L-Araf} \\
 \downarrow 2 \\
 \rightarrow 5\text{-}\alpha\text{-L-Araf}\text{-(1}\rightarrow 5\text{-}\alpha\text{-L-Araf}\text{-(1}\rightarrow 5\text{-}\alpha\text{-L-Araf}\text{-(1}\rightarrow 5\text{-}\alpha\text{-L-Araf}\text{-(1}\rightarrow \\
 \uparrow 3 \qquad \qquad \uparrow 3 \\
 \alpha\text{-L-Araf} \qquad \alpha\text{-L-Araf}
 \end{array}$$

$$\rightarrow 4)\text{-}\beta\text{-D-Galp-(1}\rightarrow 4)\text{-}\beta\text{-D-Galp-(1}\rightarrow 4)\text{-}\beta\text{-D-Galp-(1}\rightarrow 4)\text{-}\beta\text{-D-Galp-(1}\rightarrow$$
$$\begin{array}{c} \alpha\text{-L-Araf} \\ \downarrow \\ 5 \\ \alpha\text{-L-Araf} \\ \downarrow \\ 3 \\ \rightarrow 4)\text{-}\beta\text{-D-Galp}\text{-(1}\rightarrow 4)\text{-}\beta\text{-D-Galp}\text{-(1}\rightarrow 4)\text{-}\beta\text{-D-Galp}\text{-(1}\rightarrow 4)\text{-}\beta\text{-D-Galp}\text{-(1}\rightarrow \end{array}$$

β -D-Galp

↓
6

α -L-Araf-(1→3)- β -D-Galp

↓
6

→3)- β -D-Galp-(1→3)- β -D-Galp-(1→3)- β -D-Galp-(1→3)- β -D-Galp-(1→3)- β -D-Galp-

↑
6
 β -D-Galp

↑
6
 β -D-Galp

Figure 4.2 Structures of pectic polysaccharides. Homogalacturonan, rhamnogalacturonan I (RG-I), and polysaccharide side chains found on RG-I: (1→5)-α-arabinan, (1→4)-β-galactan, arabino-4-galactan, and arabino-3,6-galactan.

residues that may be methyl-esterified and acetylated to varying extents. The RG-I domain, which is usually also an abundant domain, consists of alternating α -D-GalAp and α -L-Rhap residues (Figure 4.2). The α -D-GalAp residues of RG-I may be acetylated on C(O)2 or C(O)3, or both, but it is not known if these residues may be methyl esterified (52, 53, 55). There is evidence that the HGA and XGA domains in pectin of the eudicotyledon apple (*Malus domestica*) are connected linearly with RG-I domains (56). Usually 20–80% of the Rhap residues in RG-I domains bear polysaccharide or oligosaccharide side chains rich in Araf and Galp residues, and include (1 \rightarrow 5)- α -arabinans, (1 \rightarrow 4)- β -galactans, arabino-4-galactans (57) and small proportions of arabino-3,6-galactans (Figure 4.2). Arabinans are composed of linear chains of α -L-Araf residues joined by (1 \rightarrow 5)-linkages, with side chains of single or multiple α -L-Araf residues attached through C(O)2, C(O)3 or both (Figure 4.2). Galactans consist mostly of linear (1 \rightarrow 4)- β -D-galactan chains, and arabino-4-galactans, have a (1 \rightarrow 4)- β -D-galactan backbone with side chains of α -L-Araf residues (Figure 4.2). Arabino-3,6-galactans, similar in structure to those described in the next section may also occur as side chains (55).

Ferulic acid has been shown to be ester-linked to Araf and Galp residues of RG-I in the primary walls of spinach (*Spinacia oleracea*), sugarbeet (*Beta vulgaris*), and amaranth (*Amaranthus caudatus*), all members of the Amaranthaceae (Caryophyllales) (58). Ferulic acid also occurs ester-linked to the primary walls in related “core” families of the Caryophyllales and is probably also linked to the side chains of RG-I (3, 59). In addition, a linear (1 \rightarrow 4)- β -D-galactan, chemically identical to the galactan side chain on RG-I, occurs in the lignified secondary walls of compression wood of gymnosperms, although it is not known if it forms part of an RG-I (60). In the bast fibers of flax (*Linum usitatissimum*), both short (2–3 residues) and long (up to 28 residues) (1 \rightarrow 4)- β -D-galactan chains are linked to RG-I domains (4).

A xylogalacturonan (XGA) domain occurs particularly in the walls of cells that have separated, or are about to separate, from adjacent cells, for example, in the root cap, and has single β -D-Xylp residues attached to C(O)3 of the α -D-GalAp residues of an HG backbone (61). In contrast, the RG-II domain occurs ubiquitously, but only in low concentrations. RG-II is a highly complex, low-molecular mass (\sim 5–10 kDa) domain that contains the following different monosaccharides: the hexoses D-Glcp, D- and L-Galp; the hexuronic acids D-GlcpA and D-GalpA; the pentoses L-Araf, L-Arap and 2Me-D-Xylp; the branched pentose D-Apif; the deoxyhexoses L-Fucp, L-Rhap, and 2Me-L-Fucp; and the rare sugars L-aceric acid, 2-keto-3-deoxy-D-lyxo-heptulosaric acid, and 2-keto-3-deoxy-D-manno-octulosonic acid. At least seven galacturonic acid residues form a backbone to which are attached four structurally different side chains: A, B, C, and D. A contains 2Me-D-Xylp and is an octasaccharide; B contains 2Me-L-Fucp and, depending on species, is a heptasaccharide, an octasaccharide or a nonasaccharide; C and D are disaccharides, C contains 2-keto-3-deoxy-D-manno-octulosonic acid and D contains 2-keto-3-deoxy-D-lyxo-heptulosaric acid. Except for the differences in the B-side chain, the structure of RG-II is highly conserved and occurs as a dimer cross-linked by 1:2 borate–diol esters between β -D-Apif residues on adjacent RG-II domains (62).

4.2.1.2.6 ARABINO-3,6-GALACTANS

The arabino-3,6-galactans are water-soluble polysaccharides that occur in large proportions as deposits in the lumen of the tracheids in the wood of the coniferous gymnosperm larch

(*Larix* spp.), and in smaller proportions in other softwoods (27, 63). They are composed of a backbone of (1→3)-linked β -D-Galp residues with a side chain attached at the C(O)6 position on nearly every residue. Many of the side chains are single β -D-Galp or α -L-Araf residues or the disaccharides β -L-Arap-(1→3)- α -L-Araf-(1→ or β -D-Galp-(1→6)- β -D-Galp-(1→. In some species of larch, longer, branched side chains also occur (Figure 4.2) (64). Arabino-3,6-galactans are present in spruce, pine (*Pinus sylvestris*), and larch (*Larix sibirica*) heartwoods (65–67) and carry GlcAp. Structurally similar arabino-3,6-galactans occur as covalently bound forms, as minor side chains of RG-I (see Section 4.2.1.2.6) and on hydroxyproline-rich proteins in the ubiquitous arabino-3,6-galactan-proteins (AGPs) found associated with plasma membranes and also in exudate gums of *Acacia* spp. (27, 68–70).

4.2.2 Chemistry of cell wall proteins

Proteins are ubiquitous components of plant cell walls where they may account for as much as 10% of the dry weight of the wall. The wall proteins include enzymes, wall loosening proteins (expansins) (71), and signaling molecules (AGPs) (70) but the structural proteins are quantitatively the most important. These include the glycine-rich proteins (GRP) (72, 73), the hydroxyproline-rich glycoproteins (HRGP, extensins), the proline- and hydroxyproline-rich proteins (PRP), and non-extensin proteins (70, 74–77). The HRGP are glycoproteins bearing short arabinofuranoside side chains at the Hyp residues (70). They are highly elongated molecules with extended polyproline II helix conformations imparted by the runs of Hyp residues. The HRGP, PRP, and AGP families also contain members that are chimeras having, for example, an HRGP domain fused to a domain with AGP characteristics (70). The structural proteins of the walls are thought to form an independent wall matrix phase.

4.3 Molecular associations between wall polymers

The organization of walls can be described as consisting of cellulosic microfibrils and structural proteins inserted into and reinforcing a multicomponent gel matrix, composed of stereo-irregular, non-cellulosic heteropolysaccharides (heteroglucans, heteromannans, heteroxylans, (1→3,1→4)- β -D-glucans, and pectins). In primary walls these matrix polysaccharides may interact through physical associations with one another and the cellulosic microfibrils and also through covalent associations. In secondary walls the matrix polysaccharides also interact covalently with non-carbohydrate wall components such as lignins, suberin, cutin, cutan, and proteins. The following sections describe the nature of both non-covalent and covalent associations between wall polymers.

4.3.1 Non-covalent interactions between wall polymers

In the aqueous gel matrix of the primary wall polysaccharide chains interact non-covalently to form a continuous three-dimensional network. The polysaccharide chains in the gel have two types of domains: open, hydrated, unassociated regions and regions

where the complementary conformations of two or more chains permit association over restricted segments (junction zones) (78). The forces stabilizing these junction zones are intermolecular hydrogen bonds or ionic forces. The capacity of matrix polysaccharides to form junction zones is variable and depends on the stereoregularity of the chain, determined by the monosaccharide and linkage sequence, the presence of bulky side chains and the proximity of mutually repulsive charged residues. The physical characteristics of the gel matrix will depend on the lengths and numbers of junction zones. The non-cellulosic wall polysaccharides generally have features that make them potential gel-forming polymers. They have linear backbones, are more or less soluble in water, and in contrast to the cellulosic fibrillar phase, show conformational irregularities. The non-covalent interactions of the non-cellulosic polysaccharides with one another and the surfaces of cellulosic microfibrils are important in determining the cohesivity of cell walls as is discussed in Section 4.4 – Molecular architecture of plant cell walls.

In addition to the functional strengthening of walls by non-covalent interactions between individual matrix polymers and matrix polymers and the cellulosic microfibrillar phase, these interactions are reinforced by direct covalent associations between polysaccharides, polysaccharides and lignin, polysaccharides and protein, and between proteins, as well as indirect associations between polysaccharides and polysaccharides and lignin through covalently-linked bridging molecules. The chemistry of this cross-linking is described in the following sections.

4.3.2 Covalent interactions between wall polymers

4.3.2.1 Polysaccharide–polysaccharide cross-linking.

4.3.2.1.1 DIMERIZATION OF ESTERIFIED HCA

In grasses, hydroxycinnamate esters on GAX (see Section 4.2.1.2.2) in primary and lignified secondary walls are covalently cross-linked by oxidative dimerization of HCA units on neighboring AX chains by radical coupling reactions (32, 79) (Figure 4.3a). The homo- and hetero-dehydrodimers formed involve mostly FA and sometimes SA (80); *p*CA does not appear in these dehydrodimers. Dehydrotrimers (32, 81) and a dehydrotetramer (82) of FA have also been described and could participate in cross-linking. A number of isomeric cross-linking homo- and hetero-dimers have been encountered involving both the aromatic ring and the propenoic acid side chain of the HCA (Figure 4.4) (83, 84).

4.3.2.1.2 ESTERIFIED HCA CYCLODIMERIZATION

In addition to dimerization of ester-linked HCA by oxidative coupling, cross-linking dimers are formed from FA and/or *p*CA monomers by photodimerization (83). These ester-linked cyclodimers are cyclobutane derivatives (truxillic acids and truxinic acids) (Figure 4.3b). Both the esterified HCA dimers and cyclodimers are readily released by treatment with dilute alkali (0.5 M NaOH, 20°C, 18 h) (85).

4.3.2.1.3 DIRECT ASSOCIATIONS BETWEEN POLYSACCHARIDES

Direct covalent linkages between pectic polysaccharides and xyloglucans in angiosperm walls have been reported (86–88), but the detailed chemistry of the linkage(s) is not known.

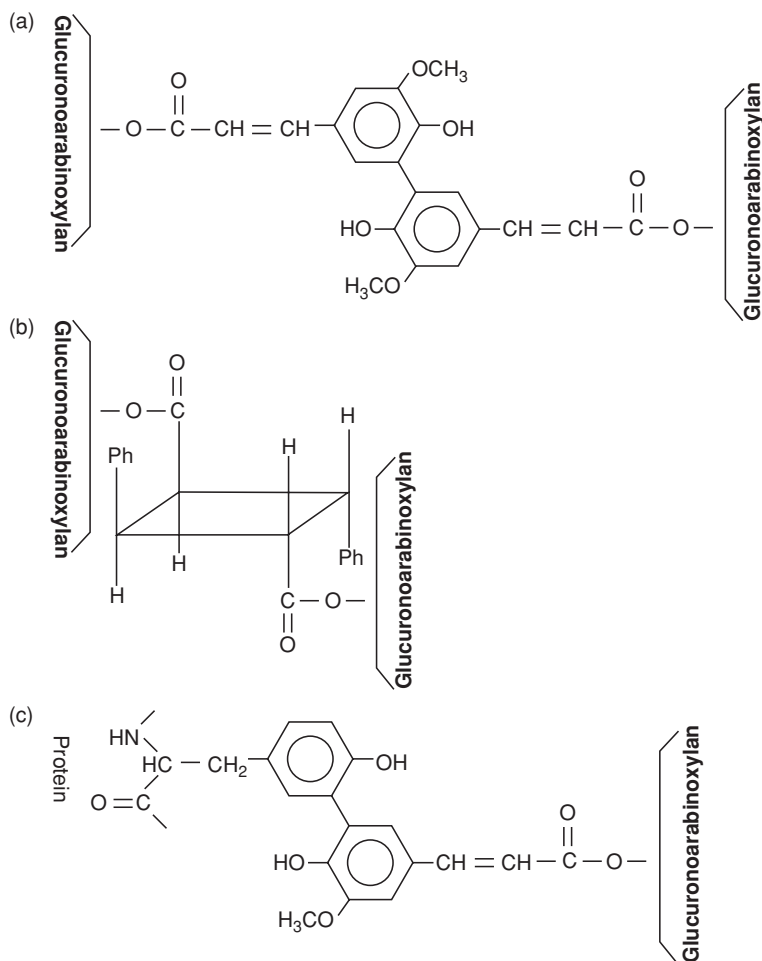


Figure 4.3 Three modes of covalent cross-linking involving feruloylated GAXs: (a) a dehydrodiferulate cross-link (5-5), (b) a ferulate cyclodimer (truxillic acid) cross-link, and (c) tyrosyl-ferulate crosslink between a protein and a feruloylated GAX. Ph = 4-Hydroxy-3-methoxy-benzene.

4.3.2.2 Polysaccharide–lignin cross-linking

Lignin–carbohydrate complexes (LCC) can be selectively isolated from lignified secondary walls of woods, for example, bald cypress (*Taxodium distichum*), birch (*Betula platyphylla*) (89), spruce (90), beech (*Fagus crenata*) and pine (*Pinus densiflora*) (91), and grasses, for example, sugarcane (*Saccharum officinarum*) bagasse (92), rice (*Oryza sativa*) (89), and wheat (*Triticum aestivum*) straws (93) by fractionation, co-extraction, and co-chromatography of native or derivatized preparations (89, 94–96). Comparative chemical (97) and physico-chemical procedures, including ¹³C-NMR (97), 2D NMR (98), and FTIR (99) have been used in the characterization of the chemistry of the lignin–polysaccharide linkages.

The carbohydrate portion of LCC from the softwood bald cypress (*Taxodium distichum*) consists of galactoglucomannan, arabino(4-*O*-methylglucurono)xylan, and

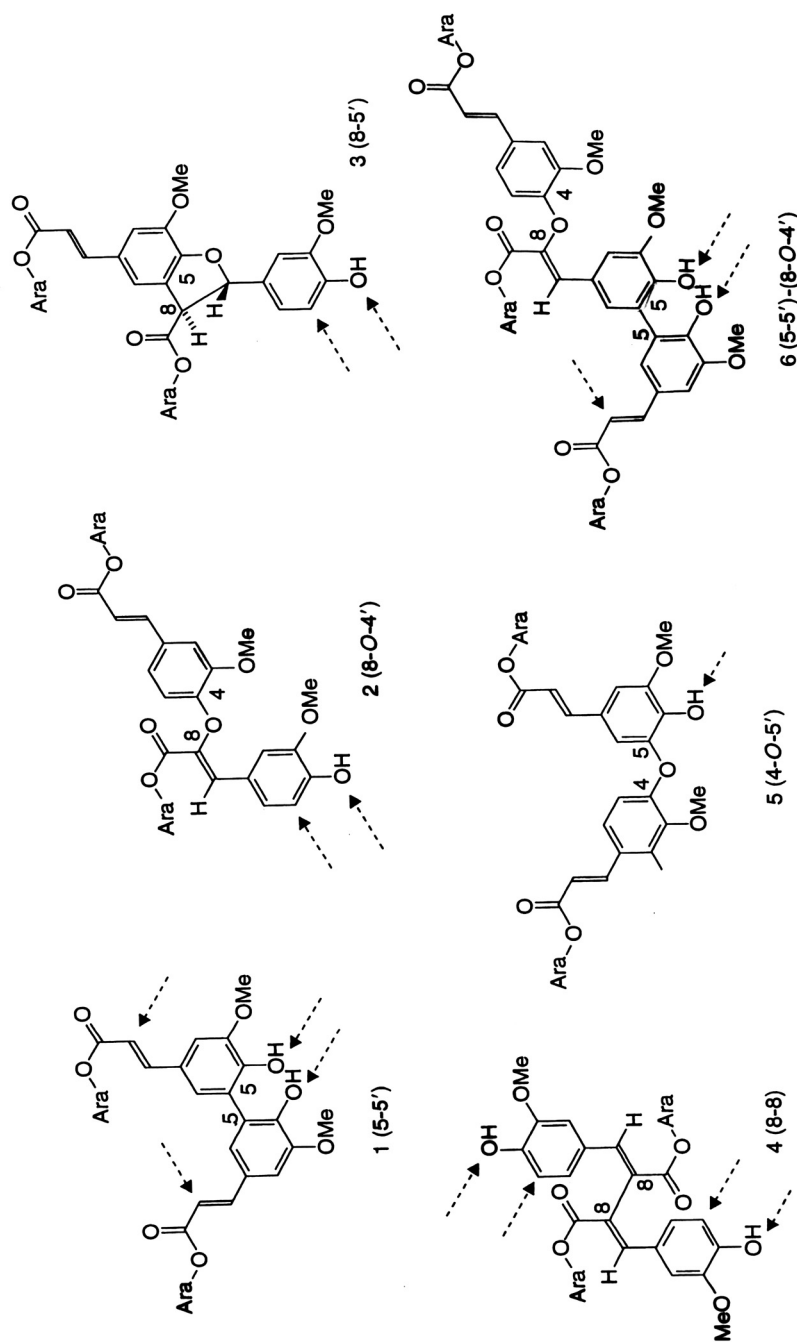


Figure 4.4 Feruloyl units on neighboring GAX chains in cell walls may be cross-linked by radical coupling into ferulate dehydrotimers (structures 1–5) or dehydrotimers (structure 6). Dotted arrows indicate potential sites for radical coupling with hydroxycinnamoyl alcohols or lignin oligomers in lignifying cell walls, resulting in cross-linking of GAXs to lignin. “Ara” is an arabinofuranosyl residue on a GAX. (Reprinted with permission, from Grabber, J.H., Hatfield, R.D., Ralph, J., Zoi, J. & Amrhein, N. (1995) Ferulate cross-linking in cell walls isolated from maize cell suspensions. *Phytochemistry*, **40**, 1077–1082, Fig. 1.)

arabinogalactan. In contrast, the LCCs of the hardwood birch (*Betula platyphylla*) and rice straw (*Oryza sativa*) are composed exclusively of 4-*O*-methylglucuronoxylans and arabino(4-*O*-methylglucurono)xylan, respectively (89). Laine and coworkers (67) found 4-linked xylan, 3,6-galactan, 4-linked galactan, and 3-linked glucan in the residual LCCs isolated from the kraft pulps of coniferous gymnosperms spruce (*P. abies*) and pine (*Pinus sylvestris*). Lawoko and coworkers (90) obtained four major LCC fractions from *P. abies* and characterized a galactoglucomannan LCC containing ~8% of the wood lignin, a glucan LCC containing ~4% of the wood lignin, a xylan–lignin–glucomannan network LCC (xylan > glucomannan) containing ~40% of the wood lignin, and a glucomannan–lignin–xylan network LCC (glucomannan > xylan) containing ~48% of the wood lignin. It was concluded that carbohydrate-free lignin, i.e., lignin without covalent bonds to carbohydrates, probably does not occur in spruce wood.

Cellulose is reported to be covalently linked to lignin in both hardwoods and softwoods (100) as judged by the identification of the polysaccharides associated with lignin in the water insoluble fraction after carboxymethyl etherification. In the softwood spruce (*Picea jezoensis*), more than half the cellulose was linked to lignin, but in the hardwood beech (*Fagus crenata*), only one-sixth. The major non-cellulosic polysaccharides of spruce wood were also covalently linked to lignin, but in beech wood the extent of their linkage to lignin was low.

4.3.2.2.1 TYPES OF COVALENT POLYSACCHARIDE–LIGNIN CROSS-LINKS

The major types of covalent linkages involved in lignin–carbohydrate associations are shown in Figure 4.5. One of these is indirect, through a bridging hydroxycinnamate molecule, the others are direct covalent linkages. These linkage types are described below.

Ester–ether cross-links. Hydroxycinnamate esters on heteroxylans and dimeric hydroxycinnamate esters bridging heteroxylan (GAX) chains, are etherified through hydroxyl(s) on lignin monomers (Figure 4.3a) and are quantitatively important in secondary walls of grasses (101), wheat (*Triticum aestivum*) (102–105) and *Phalaris aquatica* internodes (106), and maize (*Zea mays*) stover (83). The ether bond may be at the β -position of the lignin unit (102, 107–109) or possibly at the α -(benzylic)-position (110). A number of isomeric bridging homo- and hetero-dimers involved in the ester–ether bridging have been characterized (Figure 4.4) (83). The dimers involve mostly FA and dehydridiferulic acid (DDFA) and to a lesser extent SA (32, 83). *p*CA does not appear in dimeric form (32, 83). In addition, there is evidence from NMR experiments with ^{13}C -labelled Italian ryegrass (*Lolium multiflorum*) for other linkages between hydroxycinnamates and lignin monomers that are not easily broken by solvolytic analyses e.g., alkali at high temperatures (4 M NaOH, 170°C, 2 h) (103). Thus C–C, 8-*O*-4 styryl ether, and biphenyl ether coupling of ferulate and diferulates to lignins are not determined and only etherified ferulates are quantified (111–113). These may account for as little as 15% of total cross-linking (83).

Direct ester linkages. Carboxylic acid groups of uronic acids on matrix polysaccharides e.g., GAX and HGA may esterify alcoholic hydroxyls on lignin monomers (Figure 4.6). These alkali labile linkages have been reported in LCC from the woods of spruce (*Picea sitchensis*), pine (*Pinus resinosa*) and aspen (*Populus tremuloides*) and fibers from the

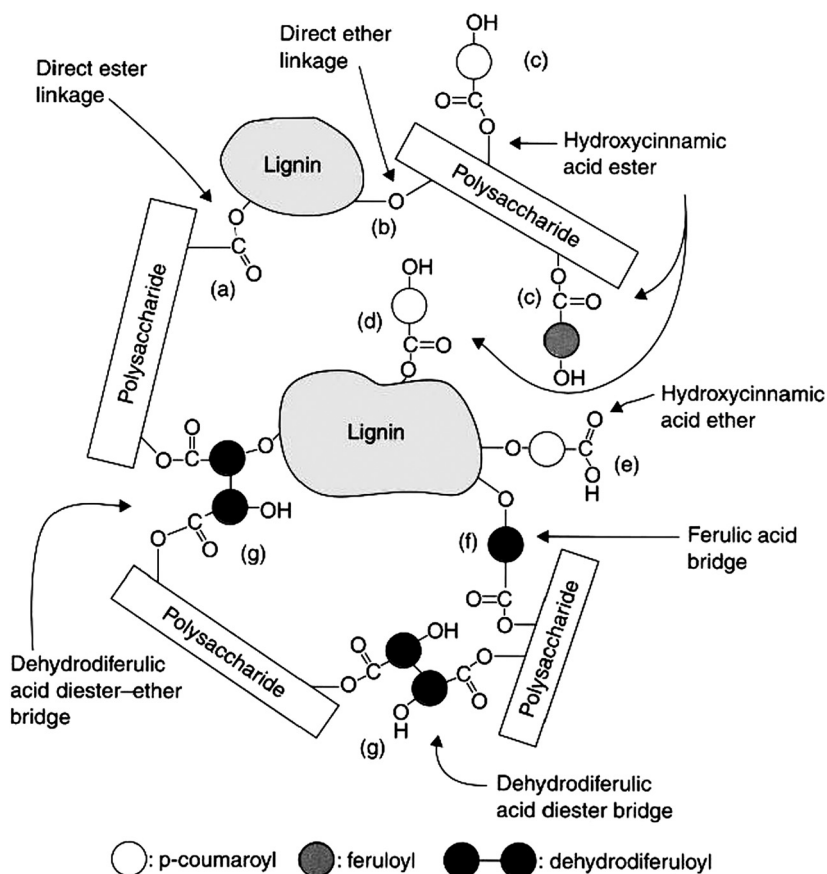


Figure 4.5 Schematic diagram showing possible covalent crosslinks between polysaccharides and lignins in secondarily thickened cell walls of commelinid monocotyledons, including grasses (○) *p*-coumaroyl, (●) feruloyl, (●●) dehydrodiferuloyl residues. (a) direct ester linkage; (b) direct ether linkage; (c) ferulic acid esterified to polysaccharide; (d) *p*-coumaric acid esterified to lignin; (e) hydroxycinnamic acid etherified to lignin; (f) ferulic acid ester-ether bridge; (g) dehydrodiferulic acid ester-ether bridge. (Reprinted with permission, from Iiyama, K., Lam, T.B.-T. & Stone, B.A. (1994b) Covalent cross-links in the cell wall. *Plant Physiology*, **104**, 315–320, Fig. 2.)

eudicotyledons jute (*Corchorus capsularis*) and mesta (*Hibiscus cannabinus*), and the commelinid monocotyledon pineapple (*Ananas comosus*) (2). Borohydride reduction of aspen LCC leads to the loss of lignin and the formation of 4-*O*-methyl glucose from 4-*O*-methylglucuronic acid residues presumably from a glucuronoxylan (114). In one case, the LCC from beech (*Fagus crenata*) wood, the entity on the lignin involved in the ester linkage has been identified as a benzyl unit using 2,3-dichloro-5,6-dicyano-1,4-benzoquinone (DDQ) and conjugate acid oxidation (91, 115). There is also direct evidence from 2D heteronuclear ^1H - ^{13}C NMR for the presence of direct ester linkages in an acetylated LCC preparation from *Eucalyptus globulus* wood (98). In this LCC no benzyl ester was detected but the γ -position of a lignin unit is esterified by a carboxyl group

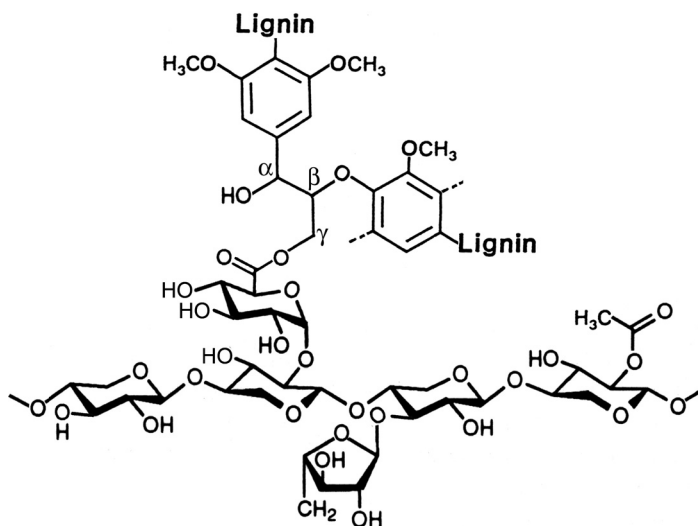


Figure 4.6 Carboxylic ester (direct ester) linkage. Structural representation of a direct ester linkage between a uronic acid carboxylic acid on a glucuronarabinoxylan and the γ -hydroxyl on the side chain of a lignin monomer.

of a uronic acid which was either GalA or GlcA or both. The extent of esterification was estimated as 9 per 100 lignin monomeric units. Pectin lyase treatment of birch (*Betula maximowiczii*) wood suggested that lignin may be esterified to HGA (116).

Benzyl ether linkages. Linkages between hydroxyls on matrix polysaccharides and lignin monomers (Figure 4.7) have been described on the basis of studies with model compounds and observations on LCC fractions (2, 83, 85). These ether linkages have been characterized by DDQ oxidation. In normal and compression woods of Japanese red pine (*Pinus densiflora*) (117), galactoglucomannan and (1 \rightarrow 4)- β -galactan were bound to the lignin

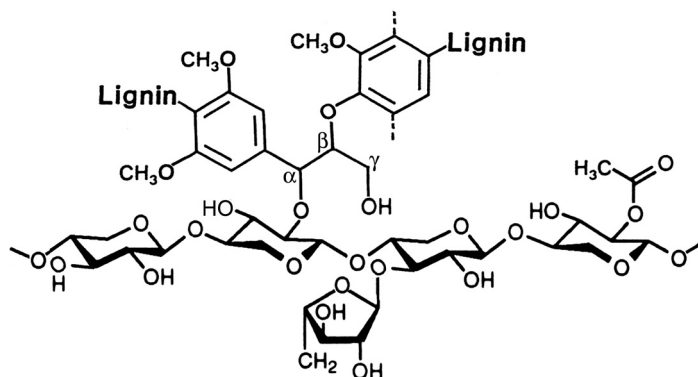


Figure 4.7 Benzyl ether linkage. Structural representation of a benzyl ether linkage between C(O)2 of a xylosyl residue of a glucuronarabinoxylan and the α -hydroxyl on the side chain of a lignin monomer.

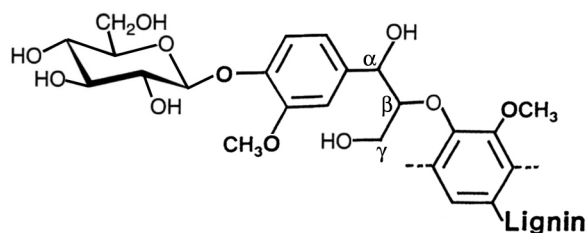


Figure 4.8 Phenyl glycoside linkage. Structural representation of a phenyl glycoside linkage between the hemiacetal hydroxyl of a β -glucose residue and the phenolic hydroxyl of a lignin monomer. Similar linkages may form between lignin and oligosaccharides and polysaccharides.

preferentially through the C(O)6 position of the hexoses and the 4-*O*-methylglucuron-arabinoxylan was bound to the lignin through C(O)2 and C(O)3 positions of xylose units (Figure 4.7). Benzyl ether linkages in *P. densiflora* LCC have also been detected by ozonolysis (118). DDQ oxidation of beech (*Fagus crenata*) wood LCC showed that xylan was ether-linked through C(O)2 and C(O)3 (117). Phenolic benzyl ether linkages, in contrast to their non-phenolic counterparts, are susceptible to alkaline hydrolysis (1 M NaOH, 100°C). They are also sensitive to acid hydrolysis (0.3 M H₂SO₄, 120°C). The lignins in the beechwood LCC are 100 times larger than those in the pinewood LCC but are less frequent (119).

Phenyl glycoside linkages. Glycosylation of lignin phenolic monomer hydroxyls or sidechain hydroxyls on lignins by monosaccharides, oligosaccharides, or polysaccharides (Figure 4.8) is a form of carbohydrate–lignin association that has experimental support (95, 120, 121). The presence of a phenyl glycoside linkage was reported in the galactoglucomannan-rich LCC from spruce (*P. abies*) wood (90). There is direct evidence from 2D heteronuclear ¹H-¹³C NMR for the presence of phenyl glycosides in an acetylated LCC preparation from *Eucalyptus globulus* wood (98). The content of phenyl glycoside moieties in the acetylated LCC was ~0.08 per monomeric lignin unit. Phenyl glycosidic linkages are acid labile but stable to alkalis.

4.3.2.3 Protein–polysaccharide cross-linking

In cell walls, proteins and polysaccharides may interact both non-covalently and covalently. Thus non-covalent associations between positively charged lysine and histidine residues on HRGP and negatively charged non-cellulosic polysaccharides, e.g., pectins and GAX have been postulated (122). Covalent intermolecular isotyrosine (phenyl ether) bridges between HRGPs have been proposed to account for the ability of acidified chlorite (123) to allow solubilization of the HRGP by cleavage of these linkages, but have not been identified, however, intramolecular isotyrosine linkages are found in HRGPs.

GAX in secondary walls of grasses may also be covalently linked to protein though dimerization of FA on GAX to tyrosine on proteins as proposed by Geissmann and Neukom (124) (Figure 4.3c). Such a mixed dimer cross-link has been isolated from an endosperm prolamin–AX complex formed during bread making (125) and has been proposed to occur in walls of lignified pericarp cells in wheat bran (126). In aleurone walls from wheat bran,

a fraction that resists digestion by hydrolases for AX and (1→3, 1→4)- β -D-glucan, contains a highly branched AX that appears to be linked to a protein, supposedly through an FA-tyrosine bridge (75).

4.3.2.4 Protein–lignin cross-linking

Covalent associations between lignin and wall proteins including GRP and HRGP have been proposed from analysis of fractions and immunocytochemical and recombinant DNA studies on secondarily thickened cell walls of mature and developing woods and legumes [reviewed by Lam and coworkers (127)] but the linkages have not been chemically characterized.

4.3.2.5 Suberin: a polyaliphatic–polyphenolic association

Some plant cells such as cottonseed hairs and the cork cells of the periderm, which forms bark, develop a suberin layer between the primary wall and the plasma membrane prior to secondary wall formation. Suberin has a polyaliphatic domain and a unique polyphenolic domain containing hydroxycinnamates (128, 129).

4.3.2.6 Cutin: an aliphatic polyester

The outer wall of the epidermal cells of stems and leaves has a multilayered cuticle, which includes a thick cuticular layer composed of cutin, a high molecular weight polyester of C16 and C18 hydroxy fatty acids which may be covalently bound to polysaccharides that form part of the underlying cell wall (130, 131). A second component of the cuticle, cutan, is a non-hydrolysable, aliphatic biopolymer believed to be composed of polyunsaturated fatty acids joined by ether bonds (132). Waxes overlie the cuticle surface.

4.3.3 Covalent cross-linking between wall polymers prevents polysaccharide utilization

In lignified secondary walls of grasses and in hardwoods and softwoods, hydrophobic lignins overlie and encrust the cellulose microfibrils and matrix polysaccharides and proteins, and are variously covalently complexed, to a greater or lesser extent, with these wall polymers. Lignins can be regarded as hydrophobic fillers that replace the water in the developing wall (133). Because water is displaced during lignification, increased hydrogen bonding is favored both between non-cellulosic polysaccharides and between these and cellulose microfibrils. Moreover, the chemical bonds between lignin and the non-cellulosic matrix components cross-link the matrix phase and the cellulose microfibrils ensuring coherence (134). Northcote (133) likened lignified walls to a synthetic glass–fiber composite, which is rigid because of the lignified matrix and has great strength because of the cellulose microfibrils. This composite has a porosity [20–50 nm in wheat straw and pericarp (135, 136)], that limits the approach of polysaccharide hydrolases to their substrates contributing to the recalcitrance of these wall types to enzymatic digestion.

Covalent cross-linking of wall polymers, in particular the lignin–polysaccharide associations, prevents extraction of matrix polysaccharides from cell walls by neutral aqueous solvents and hydrogen bond breaking reagents. Alkaline reagents, which cleave most of the

lignin–polysaccharide linkages and also solubilize some lignins (93, 137), are required. Thus, extraction of heteroxylans from lignified secondary walls of grasses can be achieved by alkali at room temperature (83), alkaline H_2O_2 (138) or oxidative degradation with chlorite (139), which however leads to polysaccharide breakdown (140). Ammonia fiber/freeze explosion effective in pretreatment of e.g., maize stover, prior to enzymatic digestion, cleaves alkali labile LCC complexes (99, 141). (See also Chapter 14, Pretreatments for Enhanced Digestibility of Feedstocks)

4.4 Molecular architecture of plant cell walls

4.4.1 Primary cell walls

4.4.1.1 Pectin-rich walls

Primary walls that are rich in pectic polysaccharides occur in the gymnosperms, eudicotyledons, non-commelinid monocotyledons, and palms (Arecaceae) (3) (see Table 4.1). The first model for this type of wall was proposed by Keegstra and coworkers (142) based on detailed analyses of the walls of cell suspension cultures of sycamore (*Acer pseudoplatanus*). This model depicted the pectic polysaccharides as being covalently linked to the xyloglucans and to the glycoprotein extensin. Hydrogen bonding between the xyloglucans and the cellulose microfibrils provided the link between the matrix complex and the cellulose. However, evidence for possible covalent linkages between pectic polysaccharides and xyloglucans was not obtained until recently (86–88). Instead, models, often referred to as “tethered or sticky network models,” were developed in which xyloglucans were postulated to form non-covalent bridges between cellulose microfibrils in walls through hydrogen bonding (143, 144).

Evidence for these non-covalent bridges was obtained using transmission electron microscopy after preparation of walls of the non-commelinid monocotyledon onion (*Allium cepa*) by fast freezing, deep etching, and rotary shadowing (145). From these studies, they developed a model in which there are two co-extensive, but independent, polymer networks: a cellulose-xyloglucan network and a pectic polysaccharide network. The first network is thought to be the main load-bearing structure of the wall, and the second is thought to determine wall porosity (Figure 4.9). A third network composed of extensin may also be present (146). Using the same technique, similar bridges between cellulose have since been observed by other researchers in primary walls of this type from a variety of species (147–149).

In these wall models, it is usually assumed that in addition to cross-linking the cellulose microfibrils, xyloglucans completely coat the surfaces of these microfibrils. However, solid-state ^{13}C NMR spectroscopy on isolated cell walls of mung bean (*Vigna radiata*) indicated a maximum of only 8% of the surface of the cellulose microfibrils had adsorbed xyloglucan (150).

The organization of the pectic polysaccharide network in these wall models is not clear but is probably in the form of a gel. Thus, the anionic HGA and RG-I have strong gel forming capabilities (53). HGA in the solid/gel state has an extended flexible conformation and may adopt a double or triple helical structure depending on the degree of hydration and the nature of the counter ion. Gel formation is due to the coordination of Ca^{2+} ions by carboxylic acid groups on adjacent chains enabling the formation of junction zones between HGA or RG-I chains. The strength of the HGA gel depends on the extent of methyl esterification of the GalAp residues.

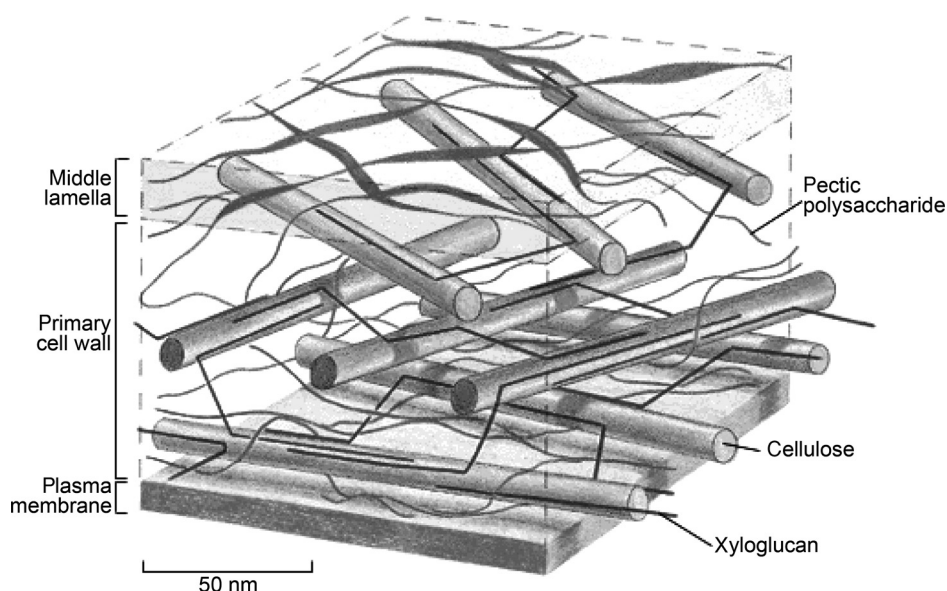


Figure 4.9 A simplified schematic representation of the spatial arrangement of polymers in a pectin-rich primary cell wall, e.g. of a parenchyma cell, as occurs in gymnosperms, eudicotyledons, non-commelinid monocotyledons, and palms (Arecaceae). The cellulosic microfibrils are embedded in a network of non-cellulosic polysaccharides [xyloglucans, pectic polysaccharides, and proteins (not shown)]. The xyloglucans are associated with the microfibril surfaces and form bridges between them. The immature primary wall may contain ~60% water but during development of the wall in some cell types (e.g., xylem fiber or tracheid), the water is replaced by lignins, which encrust the cellulosic microfibrils and non-cellulosic polysaccharides and may be covalently linked to them. (Reprinted with permission, from McCann, M.C. & Roberts, K. (1991) Architecture of the primary cell wall. In: *The Cytoskeletal Basis of Plant Growth and Form* (ed. C.W. Lloyd), pp. 109–129, Fig. 9.19, p. 126. Academic Press, London.)

The conformation of RG-I depends on the length of the alternating L-Rhap insertions into the galacturonan chain (Figure 4.2). A single L-Rhap causes a kink in the chain but two or three alternating L-Rhap residues results in a chain having a fully extended conformation comparable with HGA. RG-I side chain galactans, arabino-4-galactans etc. (see Section 4.2.1.2.4) may not seriously restrict the stereochemistry of the backbone permitting junction zone formation (53) HGA and RG-I gels are important structurally in primary cell walls and in interfaces between walls in the middle lamella, the interfacial layer between adjacent cells.

In addition to calcium bridges between HG domains, cross-linking of the pectic polysaccharide network can occur by formation of borate esters between RG-II substituents, and in the walls of “core” families of the Caryophyllales (eudicotyledons) by DDFA cross-linking of RG-I (3, 52, 62). Covalent linking of pectic polysaccharides to xyloglucans may also possibly occur (86–88).

4.4.1.2 GAX-rich walls

The primary walls of the grasses and most other commelinid monocotyledons except the palms (Arecaceae) are rich in glucuronoarabinoxylans (GAXs) (3) (Table 4.1). Examination

of the primary cell walls of the grass maize by transmission electron microscopy after preparing them by fast freeze, deep etch, and rotary shadowing showed similar bridges between the cellulose microfibrils to those described for pectin-rich primary walls [Section 4.4.1.1, (151)]. Again, two co-extensive, but independent, polymer networks have been proposed as models of the architecture of these walls (152). The bridging molecules are considered to be GAXs carrying few side chains. GAX with low degrees of substitution may interact reversibly through hydrogen bonding with surfaces of cellulosic microfibrils and may do so in the twofold helical conformation (153). With increasing degrees of substitution the affinity for microfibril surfaces decreases, presumably due to the increase in the threefold conformation which would be unfavorable for association with cellulose and additionally to increased steric hindrance to interaction imposed by the Araf substituents themselves. GAXs with low degrees of substitution with arabinosyl units, as well as $(1\rightarrow3,1\rightarrow4)$ - β -D-glucans, have been assumed to coat the cellulose microfibrils (152). However, it is possible that, as in the primary walls of mung bean (*Vigna radiata*) (150), non-cellulosic polysaccharides are adsorbed onto only a small proportion of the microfibril surfaces. Highly substituted GAXs and small proportions of pectic polysaccharides are considered to comprise the second network. The primary cell walls of all other commelinid monocotyledons (excluding the Arecaceae), including the other families that contain $(1\rightarrow3,1\rightarrow4)$ - β -D-glucans in their cell walls (17), presumably have similar wall architectures. The high molecular weight $(1\rightarrow3,1\rightarrow4)$ - β -D-glucans show high viscosities in solution and at high concentrations can be induced to form gels. Associations between $(1\rightarrow3,1\rightarrow4)$ - β -D-glucan chains are proposed to be due to junction zone formation between pairs of consecutive cellotriosyl units (154) and putatively there are also non-covalent associations between the $(1\rightarrow3, 1\rightarrow4)$ - β -D-glucan and hateroxytan components in the wall matrix (155).

4.4.2 Lignified secondary walls

Xylem fibers and tracheids are the main cell types in hardwoods and softwoods, respectively, and have thick, lignified secondary walls that have been extensively studied. In these secondary walls, the cellulose microfibrils are more densely packed and highly ordered than in primary walls, and usually occur in three layers (S_1 , S_2 , and S_3) (156). In the thinner outer (S_1) and inner (S_3) layers, the microfibrils have a cross-helical organization, whereas in the thicker middle layer (S_2), the helical organization is uniform and steeper (Figure 4.10) (157). In the S_2 wall layer, there is evidence from microscopy that aggregates of cellulose microfibrils and matrix material form alternating, concentric lamellae (158–160). In the walls of the softwood spruce (*P. abies*), the galactoglucomannans are associated more with the cellulose than the lignin, whereas with the heteroxylans the reverse is true (161, 162). These results led to a model (Figure 4.11) in which much of the galactoglucomannan is associated with the aggregates of cellulose microfibrils, and the matrix material contains lignin with associated heteroxylans and the remaining galactoglucomannans (10). This is consistent with the observations that heteromannans (glucomannans and galactoglucomannans) form oriented associations with surfaces of cellulose microfibrils since the conformation of their backbone chains is similar to cellulose (163) and are dissociated from the surfaces of cellulosic microfibrils only in strongly alkaline borate solutions [17.5% NaOH–4% borate (6 M NaOH–0.81 M H_3BO_3)] (164).

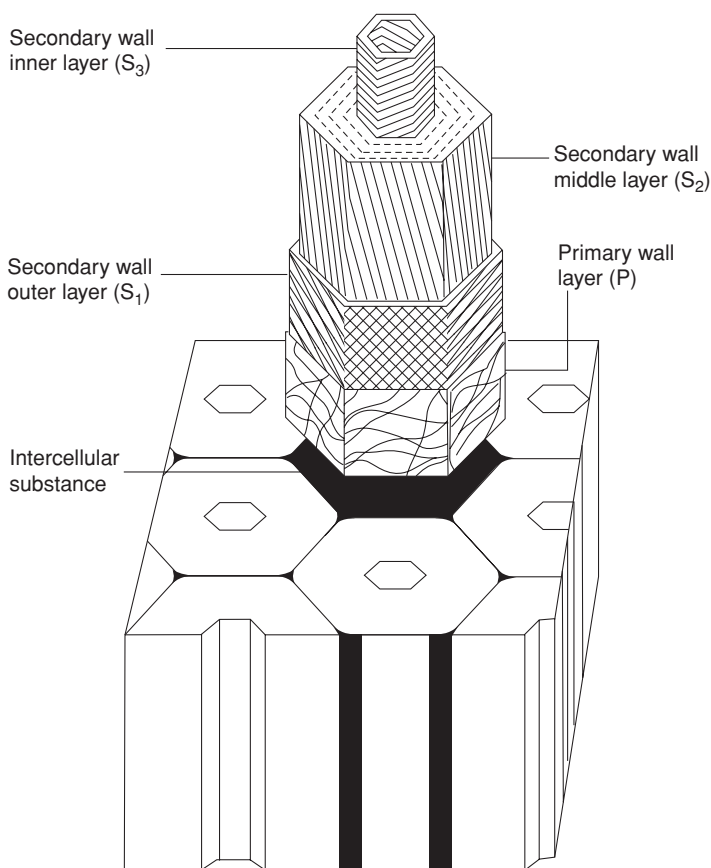


Figure 4.10 Microfibril orientation in the primary and secondary wall layers of a xylem fiber cell or a tracheid. On the inner surface of the primary wall layer (P), the microfibrils are arranged approximately transverse to the cell axis but are considerably disposed from this direction at the outer surface. In S_1 , the outermost layer (next to the primary wall layer, P) the microfibrils are usually in a flat helix (relatively transverse), whereas in the S_2 layer they are in a steep helix (relatively longitudinal). The microfibrils in the S_3 layer are again in a flat helix (more transverse in orientation). (Reproduced with permission, from Wardrop, A.B. & Bland, D.E. (1959) The process of lignification in woody plants. In: *Biochemistry of Wood* (eds. K. Kratzel & G. Billek), pp. 92–116, Fig. 3. Pergamon Press, London.)

In the xylem fibers of hardwoods, immunogold labeling indicates that heteroxylans (4-*O*-methylglucuronoxylans) with low degrees of substitution are associated with the aggregates of cellulose microfibrils, whereas heteroxylans with higher degrees of substitution are in the matrix (165). A low degree of acetylation of polysaccharides enhances their water solubility (23) and may prevent their association with one another and other polysaccharides in walls.

Although the cellulose microfibrils in the secondary walls of sclerenchyma fibers of some species of grasses are organized in a similar way to those in the xylem fibers and tracheids of hardwoods and softwoods (166), the walls of bamboo fibers have many layers and are described as polylamellate (167). As in the walls of hardwood fibers, in the lignified secondary walls of the grasses there may be populations of heteroxylans with low degrees of substitution

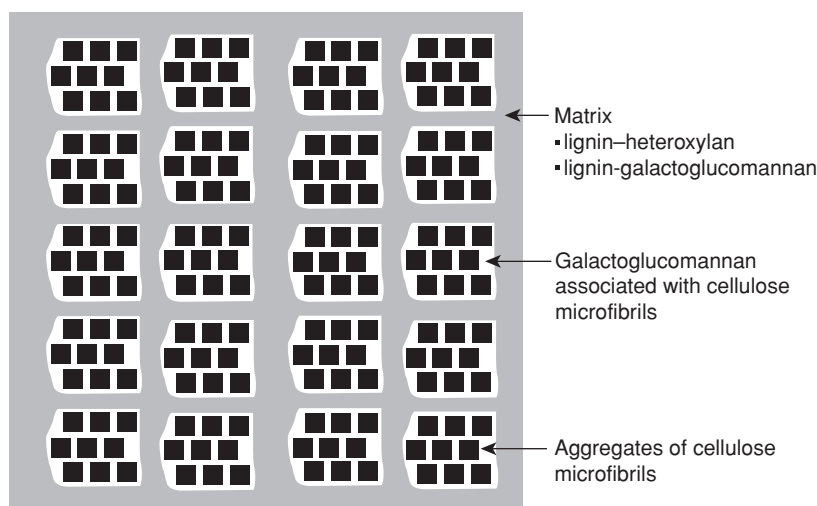


Figure 4.11 Model of a portion of the S_2 wall layer of a tracheid in the softwood spruce (*Picea abies*). Much of the galactoglucomannans are associated with aggregates of cellulose microfibrils, whereas the matrix between the aggregates contains the lignin-associated heteroxylans [arabino(4-*O*-methylglucurono)xylans] and other galactoglucomannans. (Adapted with permission, from Fahlén, J. & Salmén, L. (2005) Pore and matrix distribution in the fiber wall revealed by atomic force microscopy and image analysis. *Biomacromolecules*, **6**, 433–438, Fig. 1.)

associated with the aggregates of cellulose microfibrils and heteroxylans with higher degrees of substitution in the matrix (8, 168).

Studies of lignified secondary walls of hardwoods, softwoods, and grasses by transmission electron microscopy after preparation by fast freeze, deep etch, and rotary shadowing revealed bridges between cellulose microfibrils (147, 148, 169). Unlike the bridges in primary walls, these occur across slit-shaped pores, 8–40 nm wide, in the cell walls (148). The bridges were visible only in walls before lignification had occurred, or in lignified walls that had been delignified with acid sodium chlorite solution. Although the model proposed for the secondary wall of spruce tracheids, showed no bridges between the aggregates of cellulose microfibrils (10), similar principles used in modeling primary walls can be applied to modeling lignified secondary walls (8, 168). Thus, polysaccharides associated with the aggregates of cellulose microfibrils may also form the bridges between the aggregates.

4.5 Degradabilities of the walls of different cell types by enzymes

The degree of recalcitrance of cell wall polysaccharides to depolymerization by enzymes, or other hydrolytic reagents, depends on the ability of the enzymes to access their substrates in the walls. This is manifest at two levels. The first relates to the surface area of the wall exposed to the hydrolytic agent: the greater the extent of comminution of the feedstock, the greater the surface area exposed as the cellular organization of the plant material is disrupted (99). The second level of constraint is manifest in the organization of the polysaccharides in the

cell walls and in particular to the extent that lignin deposition and its covalent cross-linking to other wall polymers limits enzymatic degradation as discussed in Sections 4.3 & 4.4.

Information is available from studies on forages, particularly grasses and legumes, for ruminant animals on the relative degradabilities (digestibilities) of the walls of different cell types by mixtures of fungal enzymes and by rumen fluid (5, 170–174). In general, non-lignified walls are highly degradable, whereas lignified walls are much less degradable. For example, many of the parenchyma cells in stems and leaves, including mesophyll cells, have thin, non-lignified, primary walls that are highly degradable. Even moderate amounts of diferulate cross-linking between GAXs in non-lignified grass walls apparently do not impede wall degradation (175). Thus, an approximate index of the suitability of a plant as a feedstock would be the relative proportion of cells in the vegetative tissue of the mature plant that have lignified walls. The relative proportions of different cell types in a range of C₃ and C₄ grass species have been reviewed by Buxton and Redfearn (172). For example, in leaf blades of switchgrass (*Panicum virgatum*), a warm-season C₄ species, 28.9% of the cross-section area is occupied by bundle sheath cells that have moderately thick and weakly lignified walls, 35.7% by mesophyll cells with thin, non-lignified walls, and 2.9% by sclerenchyma fibers with thick, lignified walls (176). Furthermore, there may be important genetic variations in the relative proportions of the different cell types among various clones of the same species (177) that may explain variations among the clones in their total cell wall degradabilities by polysaccharide degrading enzymes, e.g., in smooth brome grass (*Bromus inermis*), a cool-season C₃ species (177). The degradability of the cell walls of internodes from 100 lines of the temperate grass (*Phalaris aquatica*) using fungal enzymes was found to be markedly dependent on the line ranging from 22 to 38% (178). In investigations of the genetic background to digestibility of smooth brome grass genotypes using fungal enzymes (179, 180), it was found that in high digestibility selections, variation in digestibility was due largely to variation in lignin concentration, whereas in low digestibility selections the influence of lignin concentration was dramatically reduced and esterified FA concentration showed a strong negative relationship with lignin concentration and appeared to be more associated with cell wall digestibility than was lignin.

During the development of forage plant organs, such as flowering stems, the overall degradability of the walls falls because some of the cell types develop lignified secondary walls (181, 182). Although this occurs in most sclerenchyma fibers and xylem tracheary elements, it occurs in only some of the parenchyma cells. Environmental conditions during growth such as temperature extremes, water deficit, nutrient stress, and shade also impact on cell wall composition and digestibility (183).

Interestingly, there are also differences between grasses and legumes in the degradabilities of their lignified walls. For example in legume stems, the lignified walls of the sclerenchyma fibers in the xylem are not degraded by rumen fluid, but in grass stems, the lignified walls of sclerenchyma fibers and parenchyma cells in grasses show significant degradability (5). However, this degradation affects only the secondary walls and not the underlying more heavily lignified primary wall and middle lamella. The cleavage by feruloyl esterases (184) of ester linkages between heteroxylans and ferulic acid linked to lignin, which occur in walls of grasses but not legumes, may be responsible for these differences.

Studies using rumen fluid and mixtures of fungal enzymes have also shown that suberized layers and cuticles on cell walls in the periderm of woody species and epidermal cells of grasses are also barriers to the access of polysaccharide hydrolases to their substrates in the wall and thus hinder their depolymerization (5, 170).

References

1. Esau, K. (1953) *Plant Anatomy*. Wiley, New York.
2. Bacic, A., Harris, P.J. & Stone, B.A. (1988) Structure and function of plant cell walls. In: *The Biochemistry of Plants: Carbohydrates*, Vol. 14 (ed. J. Preiss), pp. 297–371. Academic Press, San Diego.
3. Harris, P.J. (2005) Diversity in plant cell walls. In: *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*, Vol. (ed. R.J. Henry), pp. 201–227. CAB International Publishing, Wallingford, Oxon, UK.
4. Gorshkova, T. & Morvan, C. (2006) Secondary cell-wall assembly in flax phloem fibres: Role of galactans. *Planta*, **223**, 149–158.
5. Wilson, J.R. & Hatfield, R.D. (1997) Structural and chemical changes of cell wall types during stem development: Consequences for fibre degradation by rumen microflora. *Australian Journal of Agricultural Research*, **48**, 165–180.
6. Anon (2006) From biomass to biofuels: A road map to the energy future. In: *Biomass to Biofuels Workshop* (eds. J. Houghton, J. Ferrell & S. Weatherwax). US Department of Energy Rockville, Maryland.
7. Lewandowski, I., Scurlock, J.M.O., Linvall, E. & Christou, M. (2003) The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. *Biomass and Bioenergy*, **25**, 335–361.
8. Harris, P.J. (2005) Non-cellulosic polysaccharides in plant cell walls: Their diversity, organization and approaches to understanding their functions. In: *The Hemicelluloses Workshop 2005* (eds. K.M. Entwistle & J.C.F. Walker), pp. 13–35. The Wood Technology Research Centre, University of Canterbury, Christchurch, New Zealand.
9. Dien, B.S., Jung, H.-J.G., Vogel, K.P., Casler, M.D., Lamb, J.F.S., Iten, L., Mitchell, R.B. & Sarath, G. (2006) Chemical composition and response to dilute-acid pretreatment and enzymatic saccharification of alfalfa, reed canarygrass, and switchgrass. *Biomass and Bioenergy*, **30**, 880–891.
10. Fahlén, J. & Salmén, L. (2005) Pore and matrix distribution in the fiber wall revealed by atomic force microscopy and image analysis. *Biomacromolecules*, **6**, 433–438.
11. Davies, L.M. & Harris, P.J. (2003) Atomic force microscopy of microfibrils in primary cell walls. *Planta*, **217**, 283–289.
12. Xu, P., Donaldson, L.A., Gergely, Z.R. & Staehelin, L.A. (2007) Dual-axis electron tomography: A new approach for investigating the spatial organization of wood cellulose microfibrils. *Wood Science and Technology*, **41**, 101–116.
13. Donaldson, L. (2007) Cellulose microfibril aggregates and their size variation with cell type. *Wood Science and Technology*, **41**, 443–460.
14. Newman, R.H., Davies, L.M. & Harris, P.J. (1996) Solid-state ^{13}C nuclear magnetic resonance characterization of cellulose in the cell walls of *Arabidopsis thaliana* leaves. *Plant Physiology*, **111**, 475–485.
15. Smith, B.G., Harris, P.J., Melton, L.D. & Newman, R.H. (1998) Crystalline cellulose in hydrated primary cell walls of three monocotyledons and one dicotyledon. *Plant and Cell Physiology*, **39**, 711–720.
16. Vietor, R.J., Newman, R.H., Ha, M.A., Apperley, D.C. & Jarvis, M.C. (2002) Conformational features of crystal-surface cellulose from higher plants. *Plant Journal*, **30**, 721–731.
17. Trethewey, J.A.K., Campbell, L.M. & Harris, P.J. (2005) (1→3),(1→4)- β -D-Glucans in the cell walls of the Poales (*Sensu lato*): An immunogold labeling study using a monoclonal antibody. *American Journal of Botany*, **92**, 1669–1683.
18. Wilkie, K.C.B. (1979) The hemicelluloses of grasses and cereals. *Advances in Carbohydrate Chemistry and Biochemistry*, **36**, 215–264.

19. Smith, B.G. & Harris, P.J. (1999) The polysaccharide composition of Poales cell walls: Poaceae cell walls are not unique. *Biochemical Systematics and Ecology*, **27**, 33–53.
20. Kato, Y., Ito, S. & Yasushi, M. (2004) Study on the structures of xyloglucans using xyloglucan specific enzymes. *Trends in Glycoscience and Glycotechnology*, **16**, 393–406.
21. Hoffman, M., Jia, Z., Pena, M.J., Cash, M., Harper, A., Blackburn II, A., Darvill, A. & Albersheim, P. (2005) Structural analysis of xyloglucans in the primary cell walls of plants in the subclass *Asteridae*. *Carbohydrate Research*, **340**, 1826–1840.
22. Vincken, J.-P., York, W.S., Beldman, G. & Voragen, A.G.J. (1997) Two general branching patterns of xyloglucan, XXXG and XXGG. *Plant Physiology*, **114**, 9–13.
23. Katz, G. (1965) The location and significance of the *O*-acetyl groups in a glucomannan from Parana pine [*Araucaria angustifolia*]. *Tappi*, **48**, 34–41.
24. Mouille, G., Witucka-Wall, H., Bruyant, M.-P., Loudet, O., Pelletier, S., Rihouey, C., Lerouxel, O., Lerouge, P., Hoffte, H. & Pauly, M. (2006) Quantitative trait loci analysis of primary cell wall composition in *Arabidopsis*. *Plant Physiology*, **141**, 1035–1044.
25. Sims, I.M., Munro, S.A.L., Currie, G., Craik, D. & Bacic, A. (1996) Structural characterisation of xyloglucan secreted by suspension-cultured cells of *Nicotiana plumbaginifolia*. *Carbohydrate Research*, **293**, 147–172.
26. Ebringerová, A. & Heinze, T. (2000) Xylan and xylan derivatives – biopolymers with valuable properties. 1. Naturally occurring xylans structures, isolation procedures and properties. *Macromolecular Rapid Communications*, **21**, 542–556.
27. Whistler, R.L. & Chen, C.-C. (1991) Hemicelluloses. In: *Wood Structure and Composition* (eds. M. Lewin & I.S. Goldstein), pp. 287–319. Dekker, New York.
28. Shatalov, A.A., Evtuguin, D.V. & Neto, C.P. (1999) (2-*O*- α -D-Galactopyranosyl-4-*O*-methyl- α -D-glucurono)-D-xylan from *Eucalyptus globulus* Labill. *Carbohydrate Research*, **320**, 93–99.
29. Wilkie, K.C.B. (1985) New perspectives on non-cellulosic cell-wall polysaccharides (hemicelluloses and pectic substances) of land plants. In: *Biochemistry of Plant Cell Walls* (eds. C.T. Brett & J.R. Hillman), pp. 1–37. Cambridge University Press, Cambridge.
30. Carpita, N.C. (1996) Structure and biogenesis of the cell walls of grasses. *Annual Review of Plant Physiology and Plant Molecular Biology*, **47**, 445–476.
31. Xu, F., Sun, J.X., Geng, Z.C., Liu, C.F., Ren, J.L., Sun, R.C., Fowler, P. & Baird, M.S. (2007) Comparative study of water-soluble and alkali-soluble hemicelluloses from perennial ryegrass leaves (*Lolium perenne*). *Carbohydrate Polymers*, **67**, 56–65.
32. Ralph, J., Bunzel, M., Marita, J.M., Hatfield, R.D., Lu, F., Kim, H., Schatz, P.F., Grabber, J.H. & Steinhart, H. (2004) Peroxidase-dependent cross-linking reactions of *p*-hydroxycinnamates in plant cell walls. *Phytochemistry Reviews*, **3**, 79–96.
33. Allerdings, E., Ralph, J., Steinhart, H. & Bunzel, M. (2006) Isolation and structural identification of complex feruloylated heteroxylan side-chains from maize bran. *Phytochemistry*, **67**, 1276–1286.
34. Chesson, A., Gordon, A.H. & Lomax, J.A. (1983) Substituent groups linked by alkali-labile bonds to arabinose and xylose residues of legume, grass and cereal straw cell walls and their fate during digestion by rumen microorganisms. *Journal of the Science of Food and Agriculture*, **34**, 1330–1340.
35. Ishii, T. (1991) Acetylation at *O*-2 of arabinofuranose residues in feruloylated arabinoxylan from bamboo shoot cell-walls. *Phytochemistry*, **30**, 2317–2320.
36. Mandalari, G., Faulds, C.B., Sancho, A.I., Saija, A., Bisignano, G., LoCurto, R. & Waldron, K.W. (2005) Fractionation and characterisation of arabinoxylans from brewers' spent grain and wheat bran. *Journal of Cereal Science*, **44**, 205–212.
37. Harris, P.J., Kelderman, M.R., Kendon, M.F. & McKenzie, R.J. (1997) Monosaccharide composition of unignified cell walls of monocotyledons in relation to the occurrence of wall-bound ferulic acid. *Biochemical Systematics and Ecology*, **25**, 167–179.
38. Smith, B.G. & Harris, P.J. (1995) Polysaccharide composition of unignified cell walls of pineapple [*Ananas comosus* (L.) Merr.] fruit. *Plant Physiology*, **107**, 1399–1409.

39. Smith, B.G. & Harris, P.J. (2001) Ferulic acid is esterified to glucuronoarabinoxylans in pineapple cell walls. *Phytochemistry*, **56**, 513–519.
40. Carnachan, S.M. & Harris, P.J. (2000) Polysaccharide compositions of primary cell walls of the palms *Phoenix canariensis* and *Rhopalostylis sapida*. *Plant Physiology Biochemistry*, **38**, 699–708.
41. Darvill, J.E., McNeil, M., Darvill, A.G. & Albersheim, P. (1980) Structure of plant cell walls XI. Glucuronoarabinoxylan. A second hemicellulose in the primary cell walls of suspension-cultured sycamore cells. *Plant Physiology*, **66**, 1135–1139.
42. Timell, T.E. (1967) Recent progress in the chemistry of wood hemicelluloses. *Wood Science and Technology*, **1**, 45–70.
43. Lundqvist, J., Teleman, A., Junell, L., Zacchi, G., Dahlman, O., Tjerneld, F. & Stålbrand, H. (2002) Isolation and characterization of galactoglucomannan from spruce (*Picea abies*). *Carbohydrate Polymers*, **48**, 29–39.
44. Capek, P., Alföldi, J. & Lisková, D. (2002) An acetylated galactoglucomannan from *Picea abies* L. Karst. *Carbohydrate Research*, **337**, 1033–1037.
45. Willför, S., Sjöholm, R., Laine, C., Roslund, M., Hemming, J. & Holmbom, B. (2003) Characterisation of water-soluble galactoglucomannans from Norway spruce wood and thermomechanical pulp. *Carbohydrate Polymers*, **52**, 175–187.
46. Teleman, A., Nordström, M., Tenkanen, M., Jacobs, A. & Dahlman, O. (2003) Isolation and characterization of O-acetylated glucomannans from aspen and birch wood. *Carbohydrate Research*, **338**, 525–534.
47. Eda, S., Akiyama, Y., Kato, K., Ishizu, A. & Nakano, J. (1985) A galactoglucomannan from cell walls of suspension-cultured tobacco (*Nicotiana tabacum*) cells. *Carbohydrate Research*, **137**, 173–181.
48. Cartier, N., Chambat, G. & Josleau, J.-P. (1988) Cell wall and extracellular galactoglucomannans from suspension-cultured *Rubus fruticosus* cells. *Phytochemistry*, **27**, 1361–1364.
49. Fischer, M., Wegryzn, T.F., Hallett, I.C. & Redgwell, R.J. (1996) Chemical and structural features of kiwifruit cell walls: Comparison of fruit and suspension-cultured cells. *Carbohydrate Research*, **295**, 195–208.
50. Sims, I.M., Craik, D.J. & Bacic, A. (1997) Structural characterization of galactoglucomannan secreted by suspension-cultured cells of *Nicotiana plumbaginifolia*. *Carbohydrate Research*, **303**, 79–92.
51. Schröder, R., Nicholas, P., Vincent, S.J.F., Fischer, M., Reymond, S. & Redgwell, R.J. (2001) Purification and characterisation of a galactoglucomannan from kiwifruit (*Actinidia deliciosa*). *Carbohydrate Research*, **331**, 291–306.
52. Ridley, B.L., O'Neill, M.A. & Mohnen, D. (2001) Pectins: Structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*, **57**, 929–967.
53. Willats, W.G.T., McCartney, L., Mackie, W. & Knox, J.P. (2001) Pectin: Cell biology and prospects for functional analysis. *Plant Molecular Biology*, **47**, 9–27.
54. Willats, W.G.T., Knox, J.P. & Mikkelsen, J.D. (2006) Pectin: New insights into an old polymer are starting to gel. *Trends in Food Science and Technology*, **17**, 97–104.
55. O'Neill, M.A. & York, W.S. (2003) The composition and structure of primary plant cell walls. In: *The Plant Cell Wall* (ed. J.K.C. Rose), pp. 1–54. Blackwell, Oxford, UK.
56. Coenen, G.J., Bakx, E.J., Verhoef, R.P., Schols, H.A. & Voragen, A.G.J. (2007) Identification of the connecting linkage between homo- and xylogalacturonan and rhamnogalacturonan type I. *Carbohydrate Polymers*, **70**, 224–235.
57. Aspinall, G.O. (1980) Chemistry of cell wall polysaccharides. In: *The Biochemistry of Plants, Carbohydrates*, Vol. 14 (ed. J.E. Preiss), pp. 473–500. Academic Press, New York.
58. Bunzel, M., Ralph, J. & Steinhart, H. (2005) Association of non-starch polysaccharides and ferulic acid in grain amaranth (*Amaranthus caudatus* L.) dietary fiber. *Molecular Nutrition and Food Research*, **49**, 551–559.

59. Hartley, R.D. & Harris, P.J. (1981) Phenolic constituents of the cell walls of dicotyledons. *Biochemical Systematics and Ecology*, **9**, 189–203.
60. Timell, T.E. (1986) *Compression Wood in Gymnosperms*. Springer-Verlag, Berlin.
61. Willats, W.G.T., McCartney, L., Steele-King, C.G., Marcus, S.E., Mort, A., Huisman, M., van Alebeek, G.J., Schols, H.A., Voragen, A.G.J., Le Goff, A., Bonnin, E., Thibault, J.F. & Knox, J.P. (2004) A xylogalacturonan epitope is specifically associated with plant cell detachment. *Planta*, **218**, 673–681.
62. O'Neill, M.A., Ishii, T., Albersheim, P. & Darvill, A.G. (2004) Rhamnogalacturonan II: Structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annual Review of Plant Biology*, **55**, 109–139.
63. Grabner, M., Müller, U., Gierlinger, N. & Wimmer, R. (2005) Effects of heartwood extractives on mechanical properties of larch. *International Association of Wood Anatomists Journal*, **26**, 211–220.
64. Ponder, G.R. & Richards, G.N. (1997) Arabinogalactan from Western larch, Part III: Alkaline degradation revisited, with novel conclusions on molecular structure. *Carbohydrate Polymers*, **34**, 251–261.
65. Willför, S. & Holmbom, B. (2004) Isolation and characterisation of water soluble polysaccharides from Norway spruce and Scots pine. *Wood Science and Technology*, **38**, 173–179.
66. Willför, S., Sundberg, A., Hemming, J. & Holmbom, B. (2005) Polysaccharides in some industrially important softwood species. *Wood Science and Technology*, **39**, 245–258.
67. Laine, C., Tamminen, T. & Hortling, B. (2004) Carbohydrate structures in residual lignin-carbohydrate complexes of spruce and pine pulp. *Holzforschung*, **58**, 611–621.
68. Clarke, A.E., Anderson, R.L. & Stone, B.A. (1979) Form and function of arabinogalactans and arabino-galactan proteins. *Phytochemistry*, **18**, 521–540.
69. Fincher, G.B., Stone, B.A. & Clarke, A.E. (1983) Arabinogalactan-proteins: Structure, biosynthesis and function. *Annual Review of Plant Physiology*, **34**, 47–70.
70. Johnson, K.L., Jones, B.J., Schultz, C.J. & Bacic, A. (2003) Non-enzymic cell wall (glyco)proteins. In: *The Plant Cell Wall* (ed. J.K.C. Rose), pp. 111–154. Blackwell, Oxford.
71. Cosgrove, D.J. (2003) Expansion of the cell wall. In: *The Plant Cell Wall* (ed. J.C.K. Rose), pp. 237–263. Blackwell, Oxford.
72. Sachetto-Martins, G., Franco, L.O. & de Oliveira, D.E. (2000) Plant glycine-rich proteins: A family or just proteins with a common motif? *Biochimica et Biophysica Acta*, **1492**, 1–14.
73. Ringl, C., Keller, B. & Ryser, U. (2001) Glycine-rich proteins as structural components of plant cell walls. *Cellular and Molecular Life Sciences*, **58**, 1430–1441.
74. Jose-Estanyol, M. & Puigdomenech, P. (2000) Plant cell wall glycoproteins and their genes. *Plant Physiology and Biochemistry*, **38**, 97–108.
75. Rhodes, D.I. & Stone, B.A. (2002) Proteins in walls of wheat aleurone cells. *Journal of Cereal Science*, **36**, 83–101.
76. Cassab, G.I. & Varner, J.E. (1988) Cell wall proteins. *Annual Review of Plant Physiology and Plant Molecular Biology*, **39**, 321–353.
77. Jamet, E., Canut, H., Boudart, G. & Pont-Lezica, R.F. (2006) Cell wall proteins: A new insight through proteomics. *Trends in Plant Science*, **11**, 1360–1385.
78. Rees, D.A. (1975) Stereochemistry and binding behaviour of carbohydrate chains. In: *Biochemistry of Carbohydrates*, Vol. 5 (ed. W.J. Whelan), pp. 1–42. Butterworths, London.
79. Ralph, J., Quideau, S., Grabber, J.H. & Hatfield, R.D. (1994) Identification and synthesis of new ferulic acid dehydrodimers present in grass cell walls. *Journal of the Chemical Society, Perkin Transactions*, 3485–3498.
80. Bunzel, M., Ralph, J., Kim, H., Lu, F.C., Ralph, S.A., Marita, J.M., Hatfield, R.D. & Steinhart, H. (2003) Sinapate dehydrodimers and sinapate-ferulate heterodimers in cereal dietary fiber. *Journal of Agricultural and Food Chemistry*, **51**, 1427–1434.

81. Rouau, X., Cheynier, V., Surget, A., Gloux, D., Barron, C., Meudec, E., Louis-Montero, J. & Criton, M. (2003) A dehydrotrimer of ferulic acid from maize bran. *Phytochemistry*, **63**, 899–903.
82. Bunzel, M., Ralph, J., Brünning, P. & Steinhart, H. (2006) Structural identification of dehydrotriferulic and dehydrotetraferulic acids isolated from insoluble maize bran fiber. *Journal of Agricultural and Food Chemistry*, **54**, 6409–6418.
83. Grabber, J.H., Ralph, J., Lapierre, C. & Barriere, Y. (2004) Genetic and molecular basis of grass cell wall degradability. I. Lignin–cell wall matrix interactions. *Comptes Rendus Biologies*, **327**, 455–465.
84. Parker, M.L., Ng, A. & Waldron, K.W. (2005) The phenolic acid and polysaccharide composition of cell walls of bran layers of mature wheat (*Triticum aestivum* L. cv. Avalon) grains. *Journal of the Science of Food and Agriculture*, **85**, 2539–2547.
85. Lam, T.B.-T., Iiyama, K. & Stone, B.A. (1990) Distribution of free and combined phenolic acids in wheat internodes. *Phytochemistry*, **29**, 429–433.
86. Popper, Z.A. & Fry, S.C. (2005) Widespread occurrence of a covalent linkage between xyloglucan and acidic polysaccharides in suspension-cultured angiosperm cells. *Annals of Botany*, **96**, 91–99.
87. Brett, C.T., Baydoun, E.A.-H. & Abdel-Massih, R.M. (2005) Pectin-xyloglucan linkages in type I primary cell walls of plants. *Plant Biosystems*, **139**, 54–59.
88. Cumming, C.M., Rizkallah, H.D., McKendrick, K.A., Abel-Massih, R.M., Baydoun, E.A.-H. & Brett, C.T. (2005) Biosynthesis and cell-wall deposition of a pectin-xyloglucan complex in pea. *Planta*, **222**, 546–555.
89. Azuma, J.-I. & Koshijima, T. (1988) Lignin–carbohydrate complexes from various sources. *Methods in Enzymology*, **161**, 12–18.
90. Lawoko, M., Henriksson, G. & Gellerstedt, G. (2006) Characterisation of lignin–carbohydrate complexes (LCCs) of spruce wood (*Picea abies* L.) isolated with two methods. *Holzforschung*, **60**, 156–161.
91. Takahashi, N. & Koshijima, T. (1988) Ester linkages between lignin and glucuronoxylan in lignin–carbohydrate complex from beech (*Fagus crenata*) wood. *Wood Science and Technology*, **22**, 231–241.
92. Singh, R., Singh, S., Trimuhke, K.D., Pandare, K.V., Bastawade, K.B., Gokhale, D.V. & Varma, A.J. (2005) Lignin–carbohydrate complexes from sugarcane bagasse: Preparation, purification, and characterization. *Carbohydrate Polymers*, **62**, 57–66.
93. Durot, N., Gaudard, F. & Kurek, B. (2003) The unmasking of lignin structures in wheat straw by alkali. *Phytochemistry*, **63**, 617–623.
94. Helm, R.F. (2000) Lignin–polysaccharide interactions in woody plants. In: *Lignin: Historical, Biological, and Materials Perspectives*, pp. 161–171. American Chemical Society, New York.
95. Koshijima, T. & Watanabe, T. (2003) *Association between Lignin and Carbohydrates in Wood and Other Plant Tissues*. Springer, Berlin.
96. Watanabe, T. (2003) Microbial degradation of lignin–carbohydrate complexes. In: *Association between Lignin and Carbohydrates in Wood and Other Plant Tissues* (eds. T. Koshijima & T. Watanabe). Springer, Berlin, pp. 237–288.
97. Xie, Y., Yasuda, S., Wu, H. & Liu, H. (2000) Analysis of the structure of lignin–carbohydrate complexes by specific ^{13}C tracer method. *Journal of Wood Science*, **46**, 130–136.
98. Balakshin, H.Y., Capanema, E.A. & Chang, H.-M. (2007) MWL fraction with a high concentration of lignin–carbohydrate linkages: Isolation and 2D NMR spectroscopic analysis. *Holzforschung*, **61**, 1–7.
99. Chundawat, S.P.S., Venkatesh, B. & Dale, B.E. (2007) Effect of particle size based separation of milled corn stover on AFEX pretreatment and enzymatic digestibility. *Biotechnology and Bioengineering*, **96**, 219–231.
100. Jin, Z., Katsumata, K.S., Lam, T.B.-T. & Iiyama, K. (2006) Covalent linkages between cellulose and lignin in cell walls of coniferous and nonconiferous woods. *Biopolymers*, **83**, 103–110.

101. Atsushi, K., Azuma, J.-I. & Koshijima, T. (1984) Lignin-carbohydrate complexes and phenolic acids in bagasse. *Holzforschung*, **38**, 141–149.
102. Scalbert, A., Monties, B., Lallemand, J.-Y., Guitet, E. & Rolando, C. (1985) Ester linkage between phenolic acids and lignin fractions from wheat straw. *Phytochemistry*, **24**, 1359–1362.
103. Iiyama, K., Lam, T.B.-T. & Stone, B.A. (1990) Phenolic acid bridges between polysaccharides and lignin in wheat internodes. *Phytochemistry*, **29**, 733–737.
104. Iiyama, K., Lam, T.B.T. & Stone, B.A. (1994) Covalent cross-links in the wall. *Plant Physiology*, **104**, 315–320.
105. Beaugrand, J., Crônier, D., Thiebeau, P., Schreiber, L., Debeire, P. & Chabbert, B. (2004) Structure, chemical composition, and xylanase degradation of external layers isolated from developing wheat grain. *Journal of Cereal Science*, **52**, 7108–7117.
106. Lam, T.B.-T., Iiyama, K. & Stone, B.A. (1992) Cinnamic acid bridges between cell wall polymers in wheat and phalaris internodes. *Phytochemistry*, **31**, 1179–1183.
107. Grabber, J.H., Ralph, J. & Hatfield, R.D. (2002) Model studies of ferulate-coniferyl alcohol cross-product formation in primary maize walls: Implications for lignification. *Journal of Agricultural and Food Chemistry*, **50**, 6008–6016.
108. Jacquet, G., Pollet, B. & Lapierre, C. (1995) New ether-linked ferulic acid-coniferyl alcohol dimers identified in grass straws. *Journal of Agricultural and Food Chemistry*, **43**, 2746–2751.
109. Bunzell, M., Ralph, J., Lu, F., Hatfield, R.D. & Steinhart, H. (2004) Lignins and ferulate-coniferyl alcohol cross-coupling products in cereal grains. *Journal of Agricultural and Food Chemistry*, **52**, 6496–6502.
110. Lam, T.B.-T., Kadoya, K. & Iiyama, K. (2001) Bonding of hydroxycinnamic acids to lignin: Ferulic and *p*-coumaric acids are predominantly linked to the benzyl position of lignin, not the beta position. *Phytochemistry*, **57**, 987–992.
111. Ralph, J., Grabber, J.H. & Hatfield, R.D. (1995) Lignin-ferulate crosslinks in grasses: Active incorporation of ferulate polysaccharide esters into ryegrass lignins. *Carbohydrate Research*, **275**, 167–178.
112. Hatfield, R.D., Ralph, J. & Grabber, J.H. (1999) Cell wall structural foundations: Molecular basis for improving forage digestibilities. *Crop Science*, **39**, 27–34.
113. Grabber, J.H., Ralph, J. & Hatfield, R.D. (2000) Cross-linking of maize walls by ferulate dimerization and incorporation into lignin. *Journal of Agricultural and Food Chemistry*, **48**, 6106–6113.
114. Joseleau, J.-P. & Gancet, C. (1981) Selective degradations of the lignin-carbohydrate complex from aspen wood. *Svensk Papperstidning*, **84**, 123–127.
115. Imamura, T., Watanabe, T., Kuwahara, M. & Koshijima, T. (1994) Ester linkages between lignin and glucuronic acid in lignin-carbohydrate complexes of *Fagus crenata*. *Phytochemistry*, **37**, 1165–1173.
116. Meshitsuka, G., Lee, Z.Z., Nakano, J. & Eda, S. (1982) Studies on the nature of lignin-carbohydrate bonding. *Journal of Wood Chemistry and Technology*, **2**, 251–267.
117. Watanabe, T., Onishi, J., Yamasaki, Y., Yuko, K., Kaizu, S. & Koshijima, T. (1989) Binding site analysis of the ether linkages between lignin and hemicelluloses lignin-carbohydrate complexes by DDQ-oxidation. *Agricultural and Biological Chemistry*, **53**, 2233–2252.
118. Karlsson, O., Ikeda, T., Kishimoto, T., Magara, K., Matsumoto, Y. & Hosoya, S. (2004) Isolation of lignin-carbohydrate bonds in wood. Model experiments and preliminary application to pine wood. *Journal of Wood Science*, **50**, 142–150.
119. Takahashi, N. & Koshijima, T. (1988) Molecular properties of lignin-carbohydrate complexes from beech (*Fagus crenata*) and pine (*Pinus densiflora*). *Wood Science and Technology*, **22**, 177–189.
120. Joseleau, J.-P. & Kesraoui, R. (1986) Glycosidic bonds between lignin and carbohydrates. *Holz-forschung*, **40**, 163–168.
121. Kondo, R., Sako, T., Iimori, T. & Imamura, H. (1990) Formation of glycosidic lignin-carbohydrate complex in the dehydrogenative polymerization of coniferyl alcohol. *Mokuzai Gakkaishi*, **36**, 332–338.

122. Cooper, J.B. & Varner, J.E. (1984) Cross-linking of soluble extensin in isolated cell walls. *Plant Physiology*, **76**, 414–417.
123. O'Neill, M.A. & Selvendran, R.R. (1980) Glycoproteins from the cell wall of *Phaseolus coccineus*. *Biochemical Journal*, **187**, 53–63.
124. Geissmann, T. & Neukom, H. (1973) On the composition of the water soluble wheat flour pentosans and their oxidative gelation. *Lebensmittel-Wissenschaft und Technologie*, **6**, 59–62.
125. Piber, M. & Koehler, P. (2005) Identification of dehydro-ferulic acid-tyrosine in rye and wheat: Evidence for a covalent cross-link between arabinoxylans and proteins. *Journal of Agricultural and Food Chemistry*, **53**, 5276–5284.
126. Schooneveld-Bergmans, M.E.F., Dignum, M.J.W., Grabber, J.H., Beldman, G. & Voragen, A.G.J. (1999) Studies on the oxidative cross-linking of feruloylated arabinoxylans from wheat flour and wheat bran. *Carbohydrate Polymers*, **38**, 309–317.
127. Lam, T.B.-T., Iiyama, K. & Stone, B.A. (1990) Primary and secondary walls of grasses and other forage legume plants: Taxonomic and structural considerations. In: *Microbial and Plant Opportunities to Improve Lignocellulose Utilization by Ruminants* (eds. D.E. Atkin, L.G. Ljungdahl, J.R. Wilson & P.J. Harris), pp. 43–69. Elsevier, New York.
128. Bernards, M.A. (2002) Demystifying suberin. *Canadian Journal of Botany*, **80**, 227–240.
129. Bernards, M.A., Summerhurst, K.K. & Razem, F.A. (2004) Oxidases, peroxidases and hydrogen peroxide: The suberin connection. *Phytochemistry Reviews*, **3**, 112–126.
130. Bargel, H., Koch, K., Cerman, Z. & Neinhuis, C. (2006) Evans review no. 3: Structure–function relationships of the plant cuticle and cuticular waxes – a smart material? *Functional Plant Biology*, **33**, 893–910.
131. Stark, R.E. & Tian, S. (2006) The cutin biopolymer matrix. In: *Annual Plant Reviews* (eds. M. Riederer & C. Müller), pp. 124–144. Blackwell, London.
132. Villena, J.F., Dominguez, E., Stewart, D. & Heredia, A. (1999) Characterization and biosynthesis of nondegradable polymers in plant cuticles. *Planta*, **208**, 181–187.
133. Northcote, D.R. (1972) Chemistry of the plant cell wall. *Annual Review of Plant Physiology*, **23**, 113–132.
134. Donaldson, L.A. (2001) Lignification and lignin topochemistry – an ultrastructural view. *Phytochemistry*, **57**, 859–873.
135. Chesson, A. (1988) Lignin–polysaccharide complexes of the plant cell wall and their effect on microbial degradation in the rumen. *Animal Feed Science and Technology*, **21**, 219–228.
136. Chesson, A., Gardner, P.T. & Wood, T.J. (1997) Cell wall porosity and available surface area of wheat straw and wheat grain fractions. *Journal of the Science of Food and Agriculture*, **75**, 289–295.
137. Lapiere, C., Jouin, D. & Monties, B. (1989) On the molecular origin of the alkali solubility of Gramineae lignins. *Phytochemistry*, **29**, 1401–1403.
138. Maes, C. & Delcour, J.A. (2001) Alkaline hydrogen peroxide extraction of wheat bran non-starch polysaccharides. *Journal of Cereal Science*, **34**, 29–35.
139. Whistler, R.L. & BeMiller, J.N. (1963) Holocellulose from annual plants. In: *Methods in Carbohydrate Chemistry, Vol. III: Cellulose* (ed. R.L. Whistler), pp. 21–22. Academic Press, New York.
140. Ford, C.W. (1986) Comparative structural studies of lignin–carbohydrate complexes from *Digitaria decumbens* (pangola grass) before and after chlorite delignification. *Carbohydrate Research*, **147**, 101–117.
141. Mosier, N., Wyman, C., Dale, B.E., Elander, R., Lee, Y.Y., Holtzapple, M. & Ladisch, M. (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology*, **96**, 673–686.
142. Keegstra, K., Talmadge, K.W., Bauer, W.D. & Albersheim, P. (1973) The structure of plant cell walls. III. A model of the walls of suspension-cultured sycamore cells based on the interconnections of the macromolecular components. *Plant Physiology*, **51**, 188–196.
143. Fry, S.C. (1989) Cellulases, hemicelluloses and auxin-stimulated growth: A possible relationship. *Physiologia Plantarum*, **75**, 532–536.

144. Hayashi, T. (1989) Xyloglucans in the primary cell wall. *Annual Review of Plant Physiology and Plant Molecular Biology*, **40**, 139–168.
145. McCann, M.C., Wells, B. & Roberts, K. (1990) Direct visualisation of cross-links in the primary cell wall. *Journal of Cell Science*, **96**, 323–334.
146. McCann, M.C. & Roberts, K. (1991) Architecture of the primary cell wall. In: *The Cytoskeletal Basis of Plant Growth and Form* (ed. C.W. Lloyd), pp. 109–129. Academic Press, London.
147. Fujino, T. & Itoh, T. (1998) Changes in the three dimensional architecture of the cell wall during lignification of xylem cells in *Eucalyptus tereticornis*. *Holzforschung*, **52**, 111–116.
148. Hadrén, J., Fujino, T. & Ito, T. (1999) Changes in cell wall architecture of differentiating tracheids of *Pinus thunbergii* during lignification. *Plant and Cell Physiology*, **40**, 532–541.
149. Fujino, T., Sone, Y., Mitsuishi, Y. & Itoh, T. (2000) Characterization of cross-links between cellulose microfibrils, and their occurrence during elongation growth in pea epicotyl. *Plant and Cell Physiology*, **41**, 486–494.
150. Bootten, T.J., Harris, P.J., Melton, L.D. & Newman, R.H. (2004) Solid-state ^{13}C -NMR spectroscopy shows that the xyloglucans in the primary cell walls of mung bean (*Vigna radiata* L.) occur in different domains: A new model for xyloglucan–cellulose interactions in the cell wall. *Journal of Experimental Botany*, **55**, 571–583.
151. Satiat-Jeunemaitre, B., Martin, B. & Hawes, C. (1992) Plant cell wall architecture is revealed by rapid-freezing and deep-etching. *Protoplasma*, **167**, 33–42.
152. Carpita, N.C., Defernez, M., Findlay, K., Wells, B., Shoue, D.A., Catchpole, G., Wilson, R.H. & McCann, M.C. (2001) Cell wall architecture of elongating maize coleoptile. *Plant Physiology*, **127**, 551–565.
153. Atkins, E.D.T. (1992) Three-dimensional structure, interactions and properties of xylans. In: *Xylans and Xylanases* (eds. G.B.J. Visser, M.A. Kusters-van Someren & A.G.J. Voragen), pp. 39–50. Elsevier, Amsterdam.
154. Bohm, N. & Kulicke, W.-M. (1999) Rheological studies of barley $(1\rightarrow 3), (1\rightarrow 4)$ - β -glucan in concentrated solution: Investigation of the viscoelastic flow behavior in the sol-state. *Carbohydrate Research*, **315**, 302–311.
155. Izydorczyk, M.S. & MacGregor, A.W. (2000) Evidence of intermolecular interactions of β -glucans and arabinoxylans. *Carbohydrate Polymers*, **41**, 417–420.
156. Barnett, J.R. & Bonham, V.A. (2004) Cellulose microfibril angle in the cell wall of wood fibres. *Biological Reviews*, **79**, 461–472.
157. Wardrop, A.B. & Bland, D.E. (1959) The process of lignification in woody plants. In: *Biochemistry of Wood* (eds. K. Kratzel & G. Billek), pp. 92–116. Pergamon Press, London.
158. Ruel, K., Barnoud, F. & Goring, D.A.I. (1978) Lamellation in the S2 layer of softwood tracheids as demonstrated by scanning transmission electron microscopy. *Wood Science and Technology*, **12**, 287–291.
159. Ruel, K., Barnoud, F. & Goring, D.A.I. (1979) Ultrastructural lamellation in the S2 layer of two hardwoods and a reed. *Cellulose Chemistry and Technology*, **13**, 429–432.
160. Fahlén, J. & Salmén, L. (2002) On the lamellar structure of the tracheid cell wall. *Plant Biology*, **4**, 339–345.
161. Salmén, L. & Olsson, A.-M. (1998) Interaction between hemicelluloses, lignin and cellulose: Structure–property relationships. *Journal of Pulp and Paper Science*, **24**, 99–103.
162. Åkerholm, M. & Salmén, L. (2001) Interactions between wood polymers studied by dynamic FT-IR spectroscopy. *Polymer*, **42**, 963–969.
163. Chanzy, H. & Vuong, R. (1985) Ultrastructure and morphology of crystalline polysaccharides. In: *Polysaccharides – Topics in Structure and Morphology* (ed. E.D.T. Atkins), pp. 41–71. Macmillan, London.
164. Timell, T. (1964) Wood hemicelluloses: Part II. *Advances in Carbohydrate Chemistry*, **10**, 409–483.

165. Ruel, K., Chevalier-Billosta, V., Guillemin, F., Sierra, J.B. & Joseleau, J.-P. (2006) The wood cell wall at the ultrastructural scale – formation and topochemical organization. *Maderas: Ciencia y Tecnologia*, **8**, 107–116.
166. Juniper, B.E., Lawton, J.R. & Harris, P.J. (1981) Cellular organelles and cell-wall formation in fibres from the flowering stem of *Lolium temulentum* L. *New Phytologist*, **89**, 609–619.
167. Parameswaran, N. & Liese, W. (1976) On the fine structure of bamboo fibres. *Wood Science and Technology*, **10**, 231–246.
168. Harris, P.J. (2006) Primary and secondary plant cell walls: A comparative overview. *New Zealand Journal of Forestry Science*, **36**, 36–53.
169. Suzuki, K. & Itoh, T. (2001) The changes in cell wall architecture during lignification of bamboo *Phyllostachys aurea* Carr. *Trees*, **15**, 137–147.
170. Wilson, J.R. (1993) Organization of forage plant tissues. In: *Forage Cell Wall Structure and Digestibility* (eds. H.G. Jung, D.R. Buxton, R.D. Hatfield & J. Ralph), pp. 1–32. ASSA-CSSA-SSSA, Madison, WI.
171. Wilson, J.R., Mertens, D.R. & Hatfield, R.D. (1993) Isolates of cell types from sorghum stems. Digestion, cell wall and anatomical characteristics. *Journal of the Science of Food and Agriculture*, **63**, 407–417.
172. Buxton, D.R. & Redfearn, D.D. (1997) Plant limitations to fiber digestion. *Journal of Nutrition*, **127**, 814S–818S.
173. Hatfield, R.D., Wilson, J.R. & Mertens, D.R. (1999) Composition of cell walls isolated from cell types of grain sorghum. *Journal of the Science of Food and Agriculture*, **79**, 891–899.
174. Grabber, J.H., Panciera, M.T. & Hatfield, R.D. (2002) Chemical composition and enzymatic degradability of xylem and nonxylem walls isolated from alfalfa internodes. *Journal of Agricultural and Food Chemistry*, **50**, 2595–2600.
175. Funk, C., Braune, A., Grabber, J.H., Steinhart, H. & Bunzel, M. (2007) Moderate ferulate and diferulate levels do not impede maize cell wall degradability by human intestinal microbiota. *Journal of Agricultural and Food Chemistry*, **55**, 2418–2423.
176. Twidwell, E.K., Johnson, K.D., Patterson, J.A., Cherney, J.A. & Bracker, C.E. (1991) Degradation of switchgrass anatomical tissue by rumen microorganisms. *Crop Science*, **30**, 1321–1328.
177. Ehike, N.J. & Casler, M.D. (1985) Anatomical characteristics of smooth brome grass clones selected for in vitro dry matter digestibility. *Crop Science*, **35**, 513–517.
178. Lam, T.B.-T., Iiyama, K. & Stone, B.A. (2003) Hot alkali-labile linkages in the walls of the forage grasses *Phalaris aquatica* and *Lolium perenne* and their relation to in vitro wall digestibility. *Phytochemistry*, **64**, 603–607.
179. Jung, H.G. & Casler, M.D. (1990) Lignin concentration and composition of divergent smooth brome grass genotypes. *Crop Science*, **30**, 980–985.
180. Jung, H.G. & Casler, M.D. (1991) Relationship between lignin and esterified phenolics to fermentation of smooth brome grass fibre. *Animal Feed Science and Technology*, **32**, 63–68.
181. Lam, T.B.-T., Iiyama, K., Stone, B.A., Lee, J.A., Simpson, R.J. & Pearce, G.R. (1993) The relationship between *in vitro* enzymatic digestibility of cell walls of wheat internodes and compositional changes during maturation. *Acta Botanica Neerlandica*, **42**, 175–185.
182. Scobbie, L., Russell, W., Provan, G.J. & Chesson, A. (1993) A newly extended maize internode: A model for the study of secondary cell wall formation and consequences for digestibility. *Journal of the Science of Food and Agriculture*, **61**, 217–225.
183. Buxton, D.R. & Casler, M.D. (1993) Environmental and genetic effects on cell wall composition and digestibility. In: *Forage Cell Wall Structure and Digestibility* (eds. H.G. Jung, D.R. Buxton, R.D. Hatfield & J. Ralph), pp. 685–714. ASA, CSSA, SSSA, Madison.
184. Wong, D.W.S. (2006) Feruloyl esterase – a key enzyme in biomass degradation. *Applied Biochemistry and Biotechnology*, **133**, 87–112.

Chapter 5

Cell Wall Polysaccharide Synthesis

*Debra Mohnen, Maor Bar-Peled and
Chris Somerville*

5.1 Introduction

One of the many challenges associated with trying to understand the synthesis, structure, and function of higher plant cell walls is the obvious chemical diversity between tissue types within a plant and between plant species. For instance, an analysis of the cell wall sugar composition of different tissues in *Arabidopsis* revealed that each of the tissue types analyzed had a strikingly different composition (1). In spite of this diversity, there is substantial evidence in support of the hypothesis that most cells of all higher plants have the same six major types of polysaccharides: cellulose, xyloglucan, xylan, homogalacturonan, rhamnogalacturonan I, and rhamnogalacturonan II (2). According to this idea, variation in wall composition arises from variation in the amounts of the various polysaccharides and in variations of the structures of the various polysaccharides. Thus, in order to understand how cell walls are synthesized, much of the task may be condensed into understanding how these classes of polysaccharides are synthesized and deposited in walls. Additionally, several other polysaccharides are found in some species. These include xylogalacturonan, mannan, and mixed-linkage glucans. Lignin associated with secondary cell walls represents the other major component. Although cell wall proteoglycans are a quantitatively minor component, they may be relevant for understanding certain aspects of assembly. Similarly, callose plays an important role in synthesis of *de novo* cell walls or in specialized cell walls such as pollen tubes but is not usually a quantitatively significant component of most types of cell walls.

Plant walls are divided into two basic types, the primary and secondary wall (Figure 5.1). The primary wall is the first wall laid down in dividing and growing plant cells and it is the terminal wall in many cells in the soft parts of the plant, including the palisade cells in leaves and parenchyma cells present throughout the plant. The primary wall contains 80–90% polysaccharide and 10–20% protein. Cellulose, hemicellulose, and pectin are the main polysaccharide components in the primary wall. The most abundant hemicellulose in most primary walls is xyloglucan. However, in grasses and other commelinoid monocots glucuronoarabinoxylan is the major hemicellulose and during cell expansion in grasses, β -1,3; β 1,4-mixed linkage glucans are prevalent in the primary wall. Secondary walls are produced by specialized cells that serve a structural role such as fibers and xylem cells in vascular bundles. Secondary walls generally have less pectin, contain more cellulose and more of the hemicellulose β -1,4-xylan (often as glucuronoxylan), and are often rigidified

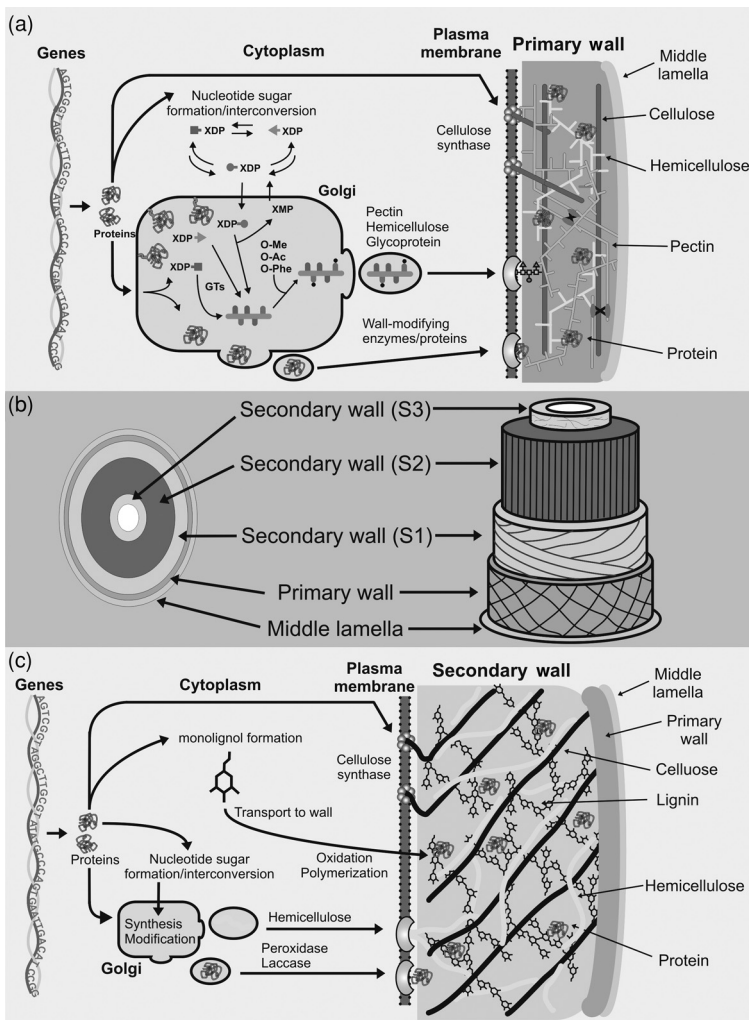


Figure 5.1 Model of plant primary and secondary walls and their synthesis. (a) Primary wall polysaccharides are synthesized at the plasma membrane (cellulose) and in the Golgi (pectin and hemicellulose) by the action of glycosyltransferases that use nucleotide-sugar substrates. (b) Some cells (e.g., xylem and fiber cells) form secondary walls internal to the primary wall. Secondary walls have increased amounts of cellulose and hemicellulose, less pectin, and are often rigidified with lignin. Secondary walls in wood tissue consist of three layers (S1, S2, and S3) that differ in cellulose microfibril orientation and chemical composition. (c) Secondary wall synthesis includes formation of cellulose microfibrils at the plasma membrane, hemicellulose within the Golgi followed by deposition in the wall, and lignin polymerization from monolignols within the wall matrix. (Figure with permission by Malcolm O'Neill, CCRC, University of Georgia)

by lignification. The synthesis of both the primary and secondary walls utilizes nucleotide-sugar substrates that are synthesized in the cytosol or in the lumen of the endoplasmic reticulum and the Golgi.

In the following overview, we have summarized the current status of knowledge about how the major classes of cell wall polymers are synthesized, modified, secreted, and assembled.

Because of the importance of *Arabidopsis* as an experimental model, much of the recent progress has been obtained using *Arabidopsis*. However, information from other species is included where it is available. The following description of cellulose synthesis is an update of a recent review (3).

5.2 Cellulose

Cellulose microfibrils are insoluble cable-like structures that are typically composed of about 36 hydrogen-bonded glucan chains each of which contains between 500 and 14 000 β -1,4-linked glucose molecules. Cellulose microfibrils comprise the core component of the cell walls that surround each cell. Studies from mutants deficient in secondary cell wall cellulose show very irregular deposition of non-cellulosic polysaccharides and lignin (4). Thus, it is apparent that cellulose is a central scaffold of cell walls.

The cellulose chains in microfibrils are parallel, and successive glucose residues are rotated 180° , forming a flat ribbon in which cellobiose is the repeating unit. The parallel chains are compatible with evidence that the chains in a microfibril are made simultaneously (3). The cellulose chains are held in a crystalline structure by hydrogen bonds and Van der Waals forces to form microfibrils. It is not yet known to what extent the "crystallization" of the nascent glucan chains into cellulose microfibrils is facilitated by proteins other than the catalytic enzyme. Jarvis (5) has shown that the two main forms of cellulose (i.e., cellulose I α and I β) can be interconverted by bending. He suggested that the sharp bend that is thought to take place when cellulose emerges from the rosette and becomes appressed to the overlying cell wall may be sufficient to induce the interconversion. Additional forms, which are primarily of interest in the context of industrial uses of cellulose, can be produced from natural cellulose by extractive treatments. For instance, in cellulose II, the chains are antiparallel – something that is unlikely to occur in native cellulose. Cellulose I is converted to cellulose II by extraction under strongly alkaline conditions.

The molecular weight of the individual glucan chains that comprise cellulose microfibrils has been difficult to determine because the extraction may lead to degradation. Analyses of secondary wall cellulose in cotton suggest a degree of polymerization (DP) of 14 000–15 000 (6). Primary wall cellulose appears to have lower molecular weight. Brown (7) reports a DP of 8000 for primary wall cellulose. However, Brett (6) reported a low molecular weight fraction of \sim 500 DP and a fraction with a DP of 2000–4000. Brett (6) suggested that the low molecular weight fraction may be chains at the surface of microfibrils whereas the high DP fraction may be chains in the microfibril interior. Since a DP of 2000 corresponds to about 1 μ m of length, the implication is that the primary wall cellulose fibrils, which are frequently observed to be much longer than 1 μ m, must be composed of chains with breaks at various locations along the fibrils. As noted below, this is compatible with genetic evidence that a cellulase is required for cellulose synthesis in both plants and bacteria (8, 9). Whatever the exact length, it is apparent that in some cells the fibrils can be extremely long relative to other types of biological macromolecules.

Based on electron micrographs, the width of cellulose fibrils varies from about 25 to 30 nm in *Valonia* and other green algae, to about 5–10 nm in most plants (10). The variation in size may indicate that cellulose microfibrils from different sources contain different numbers of chains, and it may reflect variation in the kind or amount of hemicellulose coating on the fibrils. In a study of onion primary wall by solid state NMR (10), the spectral interpretation

was consistent with the idea that the 8 nm wide microfibrils were composed of six 2-nm fibrils, each containing about 10 chains. Herth (11) estimated by electron microscopy that the microfibrils of *Spirogyra* contained 36 glucan chains. Thus, the measurements are generally consistent with the idea that each of the six globules in a rosette is composed of a number of subunits that synthesize 6–10 chains that hydrogen bond to form the 2 nm fibrils. Six of these 2 nm fibrils then bond to form the microfibrils.

The analyses of cellulose structure indicate that cellulose synthase is a highly processive enzyme, that it has many active sites that coordinately catalyze glucan polymerization, that alternating glucan units are flipped 180°, and that interspecies variation exists in the number of glucan chains per fibril, or possibly in the kind or amount of hemicellulose. What is not clear is whether the enzyme participates in facilitating the hydrogen bonding of the glucan chains or whether proximity of the glucan chains as they emerge from the enzyme is sufficient to cause formation of the highly ordered microfibrils.

Cellulose synthase can be visualized by freeze fracture of plasma membranes in vascular plants as symmetrical rosettes of six globular complexes approximately 25–30 nm in diameter. The rosettes have been shown to be cellulose synthase by immunological methods (12). The only known components of cellulose synthase in higher plants are the CESA proteins. The completion of the *Arabidopsis* genome sequence revealed that *Arabidopsis* has ten *CESA* genes that encode proteins with 64% average sequence identity (13, 14) and other species have been found to have similar numbers of CESA proteins (3). The proteins range from 985 to 1088 amino acids in length and have eight putative transmembrane (TM) domains. Two of the TM domains are near the amino terminus and the other six are clustered near the carboxyl terminus. The N-terminal region of each protein has a cysteine-rich domain with a motif that is a good fit to the consensus for a RING type zinc-finger. RING fingers have been implicated in mediating a wide variety of protein–protein interactions in complexes (15). Otherwise, the N-terminal domain is structurally heterogeneous among the ten CESAs in *Arabidopsis*. The average overall sequence identity of the amino terminal domains is 40% compared with an average overall identity of 64%.

A large “central domain” of approximately 530 amino acids lies between the two regions of transmembrane domains and is thought to be cytoplasmic. Using this feature to anchor the topology of the protein indicates that the N-terminal domain is also cytoplasmic. The central domain is highly conserved among all the CESA proteins except for an approximately 64–91 residue region of unknown significance where there is weak sequence identity. The domain contains a motif (Q/RXXRW) that is associated with bacterial cellulose synthases and other processive glycosyltransferases (16), such as chitin and hyaluronan synthases, and glucosylceramide synthase (17). Additionally, a DXXD motif and two other aspartate residues have been associated with this class of enzymes and are referred to collectively as the D,D,D,Q/RXXRW motif. Site-directed mutagenesis experiments of the chitin synthase 2 of yeast showed that the conserved aspartic acid residues and the conserved residues in the QXXRW motif are required for chitin synthase activity (18). Similarly, Saxena *et al.* (19) replaced the aspartate residues in the *A. xylinum* cellulose synthase and found that they were required for catalytic activity.

Analysis of mutants with defects in secondary wall cellulose has revealed that three separate CESA proteins are required in the same cell at the same time (20) and that the proteins physically interact (21). Thus, it appears that within a cell type there is a single type of complex containing three types of CESA subunits. A detailed summary of the properties of mutations that alter cellulose accumulation has been published recently (3). In brief, null

mutations in several of the primary wall CESA proteins are lethals, whereas others are not, presumably because of redundancy. Mutations that eliminate secondary cell wall cellulose are not lethal but impair the structural integrity of vascular cells. In addition to mutations in CESA genes, a number of other proteins have been implicated in the overall process but the role of these proteins is not understood.

5.2.1 Enzymology

Attempts to measure cellulose synthase activity *in vitro* have been problematic (3). Steady progress has been made in defining the conditions for assay and solubilization of cellulose synthase activity, although rates are still rather low (22–24). A recent study, that was carried out with large volumes of detergent-solubilized membranes from suspension cultures of *Rubus fruticosus* that facilitated structural analysis of the products (24), provided compelling evidence for synthesis of high molecular weight crystalline cellulose from UDP-glucose *in vitro*. The cellulose was visualized by electron microscopy and the properties characterized by linkage analysis and X-ray diffraction, leaving no doubt as to the identity of the *in vitro* product. The recent recovery of high levels of cellulose synthase activity from detergent extracts of membrane preparations from stationary phase suspension cultures of hybrid aspen (25) offers a new system for biomolecular studies of cellulose synthase activity.

Kudlicka and Brown (23) examined cellulose synthesized *in vitro* by electron microscopy and observed globular particles that have the same appearance as rosettes attached to the ends of the cellulose microfibrils. Similar structures were observed by Lai-Kee-Him and coworkers (24) who localized them at the non-reducing ends of the nascent cellulose fibrils. This result is in keeping with the work of Koyama and coworkers (26), where the addition of glucose units on the cellulose microfibrils from *Acetobacter aceti* was observed at the non-reducing ends of the growing ribbons. The question of the direction of chain growth remains controversial, however, since cellulose chains from *A. xylinum* were described by Han and Robyt (27) to elongate from the reducing ends. In view of evidence that β -chitin, starch, and glycogen, are polymerized from their non-reducing ends (28), and in view of the evidence from Lai-Kee and coworkers (24) it seems likely that polymerization of cellulose occurs from the non-reducing ends.

The issue of a metal requirement for catalytic activity has not yet been completely resolved by the *in vitro* studies. The addition of Mg^{++} was necessary for maximal rates of cellulose synthesis from *R. fruticosus* extracts that were solubilized with the detergent Brij 58, but inhibited activity of extracts solubilized with taurocholate (24). Activity in the absence of a divalent metal would distinguish cellulose synthase from the SGC domain proteins in which a divalent metal must bind anew at each catalytic cycle to form the nucleotide sugar-binding domain (29). Because the metal is transiently bound in SGC domain proteins, trace amounts in the assay would not be expected to support significant rates of activity. Thus, we infer that cellulose synthase is not a member of the SGC domain proteins.

The mechanism of cellulose synthesis is poorly understood. One of the persistent issues about mechanism concerns the fact that adjacent sugar residues have opposed orientations. It has been proposed that cellulose synthase has two active sites, one for each orientation

in order to facilitate the simultaneous polymerization and extrusion of the linear polymer (26, 30). The same situation applies to the processive glycosyltransferases that make chitin, hyaluronan, and heparin. Recently the first test of the two site model was reported for chitin synthase (31). These authors reasoned that if there are two UDP-GlcNAc-binding sites in close proximity, then dimeric nucleoside inhibitors should be more potent inhibitors of catalysis than the corresponding monomers. Potential bivalent inhibitors were synthesized by linking together 5'-deoxy-5'-aminouridine residues connected by ethylene glycol linkers of various lengths. Certain dimers were an order of magnitude more potent than monomeric derivatives, supporting the idea of a two-site mechanism. Conversely, UDP-chitobiose was not a substrate for chitin synthesis, mediating against the idea that an accessory protein might first condense two molecules of UDP-GlcNAc as a substrate for the synthase (32). Although these results suggest a two-site model, the CESA proteins contain only one QXXRW motif, suggesting that if two sites exist, they have distinct structural features (19).

A potentially important observation was made by Peng and coworkers (33) following inhibition of cellulose synthesis in cotton with the inhibitor CGA 325'615. They treated the resulting cell walls with cellulase with the intention of releasing cellulose synthase from nascent cellulose microfibrils. They observed that a tryptic peptide corresponding to residues 388–413 of Arabidopsis CESA1 was modified by mass amounts equivalent to the addition of 2–6 glucose residues. This seems to imply that a covalent attachment of glucan to the protein is involved in cellulose synthesis. Retaining glycosyltransferases may contain a transient covalent linkage between an Asp of the enzyme and the reducing end of the growing glycan chain (29). However, CESA proteins are considered to be members of family 2 glycosyltransferases and are proposed to function as inverting enzymes (34), which do not have such a predicted intermediate. Peng and coworkers (33) suggest that CGA 325'615 may have caused some abnormal linkage to be created.

There has been persistent interest in the concept that cellulose synthesis is initiated from a primer. Studies of the matter using bacterial synthase are controversial (9). Delmer and colleagues have suggested that sterol glucoside is a primer for cellulose synthesis (35). One line of evidence is that expression of cotton CESA1 in yeast caused formation of sterol cellodextrin from exogenously supplied sterol glucoside. Although this is intriguing, the ability to modify sterol glucoside at extremely low rates under highly artificial conditions does not mean that a primer is involved *in vivo*; many enzymes are assayed with artificial substrates that bear limited structural similarity to *in vivo* substrates. A second line of evidence is that treatment of cotton fibers with DCB reduces incorporation of radioactive glucose into sterol glucoside. Since the mode of action of DCB is not known, it could be that DCB acts by inhibiting formation of UDP-glucose or through some other indirect effect. Indeed, evidence was presented that exogenous addition of sterol glucoside could overcome the effects of DCB on *in vivo* cellulose synthesis in cotton fibers. Although the results are interesting, the demonstrations of *in vitro* cellulose synthesis did not require addition of any primer. The hypothesis should be tested by analysis of mutants of Arabidopsis with defects in the synthesis of sterol glucosides, as suggested by Peng and coworkers (35). Unfortunately, analysis of an Arabidopsis mutant with TDNA insertions in genes for the two known sterol glycosyltransferases indicated only a 40-fold reduction in sterol glucosides (Scheible and coworkers, unpublished results), which renders the absence of any apparent effect on cellulose synthesis ambiguous.

Ihara and coworkers (36) expressed the central domain of GhCESA2 in *Pichia pastoris* and found that it was soluble. It catalyzed incorporation of glucose into a product in the presence of an extract from cotton ovules but the product was not β -1,4-glucan.

5.2.2 Cellulose deposition

A distinguishing feature of plant cells is the presence of cortical microtubules adjacent to the plasma membrane (37). It has been noted since the discovery of cortical microtubules that the orientation of cortical microtubules in expanding cells is similar to that of cellulose microfibrils (38). This led to the hypothesis that the deposition of cellulose is oriented by an interaction between cellulose synthase and the microtubules, an idea that was reinforced by many observations of correlations between microtubule and microfibril organization which have been comprehensively and critically reviewed by Baskin (39). In the model of Giddings and Staehelin (40), as recast in an influential textbook (41), the movement of cellulose synthase is constrained by a close association between cortical microtubules and the plasma membrane, much like a bumper car bouncing along between rails of cortical tubulin. It is generally assumed that the energy of polymerization provides the motive force that moves the cellulose synthase complex through the membrane.

However, as noted in a recent critique of the model, there is no direct evidence for involvement of microtubules in microfibril orientation and many inconsistencies mediate against the idea (42). For instance, short treatment of *Arabidopsis* with the microtubule destabilizing drug oryzalin or the microtubule stabilizing drug taxol caused no apparent change to the orientation of cellulose microfibrils in cells that expanded during the treatment, as visualized by field emission scanning electron microscopy (43, 44). Long treatments caused changes in cellulose orientation but these may have been due to effects on the orientation of cell division. Similarly, when microtubule polymerization was impaired by shifting the temperature-sensitive *mor1-1* mutant to non-permissive temperature, cellulose microfibrils exhibited a similar pattern of deposition as in controls (45, 46).

Recently, Paredez and coworkers (47) produced a functional N-terminal YFP fusion to CESA6 that complemented the corresponding mutant in *Arabidopsis*. When expressed under the native promoter, a substantial amount of the fusion protein accumulates in the Golgi apparatus where it assembles into distinct particles that can be seen to move to the plasma membrane. This is compatible with previous evidence from electron microscopy indicating that cellulose synthase rosettes assemble in the Golgi (48). Within less than a minute of arriving in the plasma membrane, the cellulose synthase particles begin moving in linear paths at a constant rate of about 300 nm min⁻¹, somewhat slower than the rate observed by Hirai and coworkers (49) on tobacco membrane sheets. This is reminiscent of yeast chitin synthase III, in which activity is regulated by a specialized mechanism of vesicle sorting coupled with endocytic recycling (50). In this model, chitin synthase is maintained inside specialized vesicles called chitosomes (TGN/early endosome vesicles) and is transported to the specific sites of function where it becomes activated. Inactivation occurs via endocytosis. Because plant Golgi do not synthesize cellulose, it is apparent that the cellulose synthase complexes observed there are not active but that they become activated upon arrival at the plasma membrane. Rosettes have also been estimated to have only a 20 minutes lifetime in moss (51), which may suggest that they are also dissociated or endocytosed.

When viewed in cells in which the microtubules are labeled with CFP, the YFP-labeled cellulose synthase particles can be seen to move along the microtubules. Importantly, inhibition of tubulin polymerization with oryzalin rapidly leads to strong disruptions of the normal patterns of movement of the cellulose synthase particles that aggregate in patterns resembling meandering streams. Similarly, treatment of seedlings with Morlin, a novel inhibitor of microtubule treadmilling and membrane attachment, caused stalling of the cellulose synthase complexes (52). Thus, from live cell imaging it is readily apparent that microtubules exert a strong effect on the orientation of cellulose synthase movement (which presumably reflects cellulose synthesis) (47). However, Paredez and coworkers (47) observed that after relatively long periods of oryzalin treatment, when most or all of the cortical microtubules have depolymerized, the cellulose synthase particles resume movement in relatively straight parallel paths. The rigidity of cellulose probably explains why no guidance is necessary to ensure that cellulose synthase moves in relatively straight lines. It is not clear what orients the pattern of deposition in these cells but models for the formation of oriented patterns of cellulose based on geometric considerations have been proposed (53) and may be testable in these experimental materials. These observations suggest that both sides of the microtubule–microfibril alignment debate are correct and that the discrepancies and inconsistencies between experiments reflect the limitations of using static imaging methods and different treatment times and conditions. The availability of the new imaging tools outlined here should facilitate a resolution of the matter.

Alignment of GFP-labeled cellulose synthase with microtubules was previously reported by Gardiner and coworkers (54), who used an N-terminal fusion of GFP to the xylem-specific CESA7 (*irx3*) protein. Because of difficulties viewing the vascular tissues by confocal microscopy, the images of this GFP:CESA7 construct are difficult to discern. However, it appears that the distribution of fluorescence is not uniform and there are bands of fluorescence that are perpendicular to the long axis of the cells. Attempts to colocalize tubulin with CESA7 using immunofluorescence methods (54) indicate a similar pattern. However, the resolution of the images was not high enough to provide a critical analysis. Treatment with the microtubule assembly inhibitor, oryzalin, rapidly reduced the banding pattern. Given the technical limitations of working with xylem-localized markers, the observations of Gardiner and coworkers (54) appear to be entirely consistent with the more recent work of Paredez and coworkers (47).

A surprising twist to the microtubule–cellulose synthase story was the observation that in tobacco protoplasts, inhibition of cellulose synthase activity prevented the development of oriented microtubule arrays (55). These data are consistent with the hypothesis that cellulose microfibrils or cellulose synthase, directly or indirectly, provide spatial cues for cortical microtubule organization. Similarly, microtubule organization in spruce pollen tubes was altered by isoxaben (56), and the orientation of microtubules in *Arabidopsis* root epidermal cells was disrupted by DCB (46).

5.2.3 Regulation of cellulose synthesis

In bacteria, cellulose synthase appears to be constitutively produced and is activated by the regulatory molecule Bis-(3'-5')-cyclic dimeric guanosine monophosphate (9, 57). C-di-GMP has not been found in plants but cotton fibers were reported to have a binding

protein (58). However, comparison of the sequence of the apparent binding protein with the Arabidopsis proteome indicates that the putative binding protein is α -tubulin or something that copurified with it.

As noted above, recent results suggest that plant cellulose synthase is activated by a process associated with secretion. In principle, a plasma membrane-localized kinase or phosphatase could alter the activation state of cellulose synthase following transfer from the Golgi, providing a mechanism for keeping it inactive in the Golgi but rapidly activating it upon arrival in the plasma membrane. A proteomics survey of plasma membrane phosphoproteins revealed that CESA1, CESA3, and CESA5 proteins were phosphorylated at a number of sites, and several of the peptides had more than one residue phosphorylated (59). The sites were clustered in the N-terminal domain and in the hypervariable region of the central domain (3). Analysis of the CESA7 protein by mass spectrometry showed that two serine residues in the hypervariable region are phosphorylated (60). Cell extracts catalyzed phosphorylation of these residues and the phosphorylated polypeptide region was rapidly degraded by a proteasome-dependent pathway, leading to the suggestion that phosphorylation may regulate protein turnover. The KOR protein, which is required for cellulose synthase activity *in vivo*, also had at least two phosphorylated peptides.

During cell expansion, cellulose synthesis is a major consumer of fixed carbon. Thus, it seems likely that whatever regulates cellulose synthesis is coordinated with other aspects of primary carbon metabolism. In plants, UDP-glucose is thought to be largely synthesized by sucrose synthase (SuSy) (61). Amor and coworkers (62) observed a form of SuSy that was associated with the plasma membranes. They also observed that sucrose supported much higher rates of cellulose synthesis by extracts from developing cotton fibers than UDP-glucose and that sucrose synthase is very strongly upregulated in cotton fibers at the onset of fiber elongation. Haigler and coworkers (61) have presented an extensive review of the hypothesis that SuSy might channel UDP-glucose to cellulose synthesis. This is an attractive idea but direct evidence is lacking. Arabidopsis has six SuSy genes but no two isoforms have the same pattern of expression (63). Mutant plants lacking individual isoforms, or double mutants of closely related isoforms, had no alteration in cellulose content. Thus, these studies did not provide support for the idea that SuSy is an important factor in controlling cellulose synthesis. By contrast, transgenic suppression of several SuSy genes in developing cotton fibers prevented formation of fiber cells (64). The effect was more profound than could be attributed solely to an inhibition of cellulose synthesis, obscuring a mechanistic interpretation of the effects. Increased expression of various forms of SuSy in transgenic tobacco plants did not result in increased cellulose per cell, suggesting that UDP-glucose is not the limiting factor in cellulose accumulation in that system (65).

Analysis of the steady-state level of mRNA in major tissues of Arabidopsis with gene chips showed that the *CESA1*, 2, 3, 5, and 6 genes are expressed in all tissues at moderately high levels that differ by about fourfold at most (66). Similar results can be compiled from the large number of public microarray datasets that are now available for Arabidopsis from sites such as Genevestigator (67). As noted below, *CESA1*, 2, 3, and 6 have been implicated in primary wall synthesis by mutant analysis. Analyses of expression of *CESA* genes in Arabidopsis embryos revealed that *CESA1*, 2, 3, and 9 are the only *CESAs* expressed there (68). Thus, following the nomenclature of Burton and coworkers (69) *CESA1*, 2, 3, 5, 6, 9 are probably involved in primary wall synthesis and are referred to as Group-I *CESAs*. By contrast *CESA4*, 7, 8 are mostly or only expressed in tissues such as stems where secondary cell walls are found

and are designated Group-II (21, 66). *CESA4* promoter:GUS expression studies confirmed that the *CESA4* gene was mostly or only expressed in the vascular tissues (13). Similarly, immunological staining of tissue prints with antibodies against *CESA7* and *CESA8* showed that the corresponding genes were only expressed in the xylem and interfascicular region (70).

Maize has at least 12 *CESA* genes (71). PCR analysis of transcript levels of six of the genes in various tissues indicated that all of the genes were expressed in all of the tissues examined (13). Analysis of eight of the maize genes by massively parallel signature sequencing indicated that the levels of several of the *CESA* genes varied from one tissue type to another, but no conclusions were reached concerning functional specialization (72). A subsequent analysis that included three additional genes resulted in the identification of three genes that were specifically associated with secondary cell wall formation (71). Thus, maize also shows evidence for specialization of primary and secondary cell wall syntheses.

Quantitative information about the relative levels of expression of the Arabidopsis *CESA* genes is lacking because the gene chips used for most studies have not been calibrated for the various *CESA* genes. By contrast, Burton and coworkers (69) used quantitative PCR to measure the expression of the eight known barley *CESA* genes. They observed that the *CESA* genes could be grouped into two expression patterns (i.e., Group I and II) that were generally consistent with roles in primary and secondary wall synthesis. Additionally, they observed that there were large differences in the relative abundance of transcripts for the various members of a *CESA* group. If the *CESA* genes are translated with similar efficiency, this observation would suggest that the various *CESA* proteins are not present in identical amounts in the *CESA* complexes.

Consistent with genetic evidence that at least three *CESA* proteins are required to produce a functional cellulose synthase complex, correlation analysis of public and private DNA chip datasets revealed that expression of the Arabidopsis *CESA4*, 7, 8 gene were indeed very highly correlated (73, 74). The expression of a number of other genes was also very highly correlated with these genes and insertion mutations in several of these genes resulted in cellulose deficient phenotypes. Mutations in some highly correlated genes did not result in obvious effects on cellulose synthases but resulted in other defects in secondary wall synthesis. Thus, the evidence is compatible with the idea that the *CESA* genes that participate in secondary wall synthesis are under developmental control along with other genes required for secondary wall synthesis. The *CESA* genes implicated in primary wall synthesis were less highly correlated. This is consistent with the observation that there are more than three *CESA* genes associated with primary wall synthesis. This presumably indicates that some of the Group-I *CESAs* are functionally redundant and, therefore, their expression may vary from one tissue to another for unknown reasons. For instance, as noted above, *CESA9* appears to be specifically expressed in embryos.

There is sparse evidence suggesting that cellulose synthesis may be regulated in response to stimuli other than developmental programs. Transgenic trees in which 4-coumarate:coenzyme A ligase expression was reduced by expression of an antisense gene exhibited up to a 45% reduction of lignin and a 15% increase in cellulose (75). However, the apparent increase in cellulose may have been due to a decrease in total mass caused by the reduced lignin content. Conversely, antisense-mediated reduction in expression of an α -expansin in petunia caused a significant reduction in cellulose accumulation in petals (76). According to current theories of expansin action (77) this presumably reflects an

indirect effect from a defect in cell expansion. The properties of this mutant raise the possibility that many or all mutants with defects in cell expansion may have reduced cellulose content due to some form of feedback regulation of cellulose synthesis.

5.3 Hemicellulose

Hemicellulose is an operational term that refers to polysaccharides that are extracted from cell walls by dilute alkali. Xylans are typically extracted by 4% KOH whereas xyloglucans may require 24% KOH. The extraction conditions are thought to dissociate hydrogen bonding between hemicellulose and cellulose. By this definition, hemicelluloses constitute between a quarter and a third of the mass of many types of cell walls. The composition of the hemicellulose fraction varies from one cell type to another and among plant species. The four types of polymers that comprise most types of hemicellulose are xyloglucan, xylans, mixed-linkage glucans, and mannans such as glucomannan, galactomannan, or galactoglucomannan. Xyloglucan is defined by a β -1,4-glucan backbone that is highly substituted with xylose-containing side chains that may also contain galactose, arabinose, or fucose. Xylans contain a β -1,4-xylose backbone which may contain arabinan and glucuronic acid side chains. Mixed linkage glucans are composed of glucose residues linked both β -1,3 and β -1,4. Mannans have a β -1,4-linked mannose backbone or, in the case of glucomannans, may contain both glucose and mannose linked β -1,4. Konjac glucomannan, which is an important dietary fiber, contains glucose:mannose in a ratio of 1:1.6, indicating that the residues do not simply alternate. By contrast, galactomannans have a β -1,4-mannose backbones with α -1,6-galactose branches. The ratio of mannose:galactose varies from about 1:1 in fenugreek gum to about 4:1 in locust bean gum.

Several of the proteins involved in the synthesis of the various hemicelluloses polymers have recently been identified. Somewhat interestingly, it appears that the backbones of several of the hemicellulose polysaccharides are synthesized by a family of related enzymes that were originally termed as “cellulose synthase like” (CSL) because of weak sequence similarity to the CESA proteins that comprise cellulose synthase.

5.3.1 Mannan

Polysaccharides with β -1,4-mannan and β -1,4-glucomannan backbones are abundant constituents of the wood of gymnosperms (78, 79) the cell walls of certain algae (80), and are present in lower amounts in many other species (81). Mannans also serve as carbohydrate reserves in a variety of plant species (78, 82). Several groups have biochemically characterized glucomannan synthase activities from a variety of plant species (78, 79, 82, 83). The enzymes were shown to use GDP-mannose and GDP-glucose as substrates and to produce polymers with varying ratios of the two sugars, depending on the ratio of the sugar nucleotides (83). Recently, transglycosylase enzymes that modify the architecture of mannan polysaccharides in plant cell walls have also been discovered (84).

Mannan synthase is encoded by genes of the CSLA gene family (85). Expression of a cDNA from guar in soybean cells led to the synthesis of a β -1,4-mannan. Similarly, expression of three Arabidopsis CSLA proteins in *Drosophila* cells resulted in proteins that catalyzed synthesis of mannan from GDP-mannose (86). Similarly, several poplar orthologs expressed in *Drosophila* cells exhibited glucomannan synthase activity (87). Functional analysis of

CSLA genes from diverse species is consistent with the hypothesis that the function of the CSLA genes is conserved in all plants (81).

As noted previously in studies of impure enzymes (83), when provided with both GDP-glucose and GDP-mannose, the enzymes produce mixed linkage mannans. One of the proteins produced glucan when provided only with GDP-glucose. These studies suggest that no primer is required to initiate mannan synthesis. The ability of the CSLA enzymes to accept GDP-glucose or GDP-mannose is compatible with earlier suggestions that the ratio of mannose:glucose in glucomannans may be controlled by regulating the availability of the two sugar nucleotides. The observation that GDP-glucose is a substrate also may explain old observations from in vitro polysaccharide synthesis experiments that had initially been interpreted as evidence that GDP-glucose was the substrate for cellulose synthesis.

Genes encoding the galactomannan galactosyltransferase responsible for attaching galactan side chains have also been identified following purification of the enzyme from fenugreek (88). The α -1,6GalT galactosyltransferase cDNA encodes a 51282 Da protein, with a single transmembrane alpha helix near the N terminus. The protein has been functionally expressed in the yeast *Pichia pastoris* and is active when the membrane-spanning domain is removed. Thus, presumably the membrane-spanning domain is required only to localize the protein to the Golgi apparatus where mannan is thought to be synthesized. The degree of substitution when UDP-galactose is available is variable and appears to be a stochastic process controlled both by enzyme specificity and the levels of α 1,6GalT activity (89). Eight *Arabidopsis* gene sequences are very similar to the α -GalT from fenugreek (88).

In *Arabidopsis*, mannans have been localized not only in thickened secondary cell walls of xylem elements, xylem parenchyma, and interfascicular fibers, but also in the thickened walls of the epidermal cell of leaves and stems and, to a lesser extent, in most other cell types (90). Analyses of *Arabidopsis* mutants containing a transposon insertion in exon seven of the *CsIA7* gene showed that disruption of this gene results in defective pollen tube growth and disruption of embryonic development (91). Mutants (called *rat4*) containing a T-DNA insertion in the 3' untranslated region of the *AtCsla9* gene display resistance to *Agrobacterium tumefaciens* transformation, apparently caused by decreased binding of bacterial cells to roots. *AtCsla9* promotor-GUS fusions indicated that this gene is expressed in a variety of *Arabidopsis* tissues, including lateral roots and the elongation zone, where the root is most susceptible to *Agrobacterium* transformation (92). In both mutant studies, the authors suggested that the mutant phenotypes resulted from alterations in polysaccharide content; however, in neither case was such a defect demonstrated.

A GDP-mannose transporter that is localized to the Golgi has been characterized (93).

5.3.2 Xyloglucan

Xyloglucans exist as cell wall components in most species and as storage polymers in seeds of some species (94). Xyloglucan (XG) comprises 20–25% of primary walls of dicots but graminaceous monocots typically contain much less. XG is defined by a β -1,4-glucan backbone in which many glucosyl residues contain α -1,6-linked xylose branches. Xyloglucan from pea had an average molecular mass of 330 kDa representing a backbone of about 1100 glucose residues of about 500 nm in length (95). In many species the xylose residues are further substituted with β 1,2-linked galactose which may in turn be linked at the 2-position

Table 5.1 Common elements of single letter code for xyloglucan structure

Code	Structure represented
G	β -D-Glc p [*] - ^a
X	α -D-Xylp-(1→6)- β -D-Glc p [*] -
L	β -D-Galp-(1→2)- α -D-Xylp-(1→6)- β -D-Glc p [*] -
F	α -L-Fucp-(1→2)- β -D-Galp-(1→2)- α -D-Xylp-(1→6)- β -D-Glc p [*] -

^a*- D-Glucose in chain or at reducing terminus. See Fry and coworkers (97) for details.

to α -L-fucose or arabinose (94). X-ray fiber diffraction studies of tamarind XG indicated a twofold helix similar to cellulose (96). A single letter code has been developed to describe the structure of xyloglucans (Table 5.1) (97).

Most species have an XXXG type of XG. However, members of *Poacea* and *Solanaceae* have an XXGG type in which a pair of arabinose residues replace fucose (98, 99). In most monocots XG contains less xylose and galactose and does not contain terminal fucose. The structure and molecular distribution of the side chains varies in different plant tissues and species (100–102).

XG may be extensively acetylated (103). In Sycamore cells, the O-2-linked- β -D-galactosyl residue of the nonasaccharide was found to be the dominant site of O-acetyl substitution in XG. Both mono- O-acetylated and di- O-acetylated β -D-galactosyl residues were detected. The degree of O-acetylation of the β -D-galactosyl residue was estimated to be 55–60% at O-6, 15–20% at O-4, and 20–25% at O-3. Approximately, 50% of the β -D-galactosyl residues were mono- O-acetylated, 25–30% were di- O-acetylated, and 20% were not acetylated. In tomato (*Lycopersicon esculentum*), O-acetyl substituents were located at O-6 of the unbranched backbone β -D-Glcp residues, O-6 of the terminal β -D-Galp residue, and/or at O-5 of the terminal α -L-Arap residues (104). Acetylation of XG does not affect the degree to which XG hydrogen bonds to cellulose in vitro (100) and the role of acetylation is unknown. Similarly, the enzymes that acetylate XG are unknown. O-acetylation of galactose residues was considerably reduced in Fuc-deficient mutants (*atfut1*, *murl1*, and *mur2*) that synthesize XG containing little or no Fuc (105). These results suggest that fucosylated XG is a suitable substrate for at least one O-acetyltransferase in Arabidopsis.

Immunoelectron microscopy using antibodies against XG indicates that XG is localized to the cellulose-containing region of the cell wall (106). Hayashi (94) proposed that XG does not have covalent cross-links to other components or if there are links they must be alkali-labile linkages such as O-esters. However, Thompson and Fry (107) have observed cross-links between XG and pectins in Rose cells. Brett and coworkers (108) have also observed such cross-links and found that they form in the Golgi. Feruloyl esters of XG have also been observed in maize cell cultures (109).

Pure XG binds to cellulose in vitro in a pH-dependent manner (110). Levy *et al.* (111) have presented evidence that the structure of the XG side branches may facilitate the binding of XG to cellulose. Native XG-cellulose complexes contain higher ratios than can be obtained in vitro, suggesting that XG may be intercalated into the cellulose microfibrils (110). Also, mild alkali does not completely dissociate the complex and concentrated alkali (e.g., 4M KOH) is required to completely extract XG. The proposed function of XG binding to cellulose is to prevent aggregation of cellulose fibrils (110) but because single strands of XG may be

hydrogen bonded to different cellulose microfibrils (112–114); it may also provide some degree of crosslinking. Thus, XG hydrolysis may be required for growth. Based on the combined chemical and cytological evidence, Pauly and coworkers (100) have developed a model for the cellulose/XG network that posits that XG can have three configurations; hydrogen bonded to the surface of cellulose, cross-linked, and embedded within the microfibril. They propose that the cross-links are the domain that is accessible to enzymes such as xyloglucan endotransglycosylases that are thought to play a role in cell wall expansion. This is supported by observations of XET-mediated incorporation of fluorescent XG fragments into XG in expanding cell walls (115). Pauly and coworkers (100) also note that it is not clear to what extent the various XG structures participate in determining the nature of the XG-cellulose association.

The first progress in defining the genes involved in XG synthesis was the identification of the fucosyltransferase that adds the terminal fucose to XG side chains. A 60-kDa fucosyltransferase (FTase) that adds this residue was purified from pea epicotyls (116). Peptide sequence information from the pea FTase allowed the cloning of a homologous gene, AtFUT1, from Arabidopsis. AtFUT1 expressed in mammalian COS cells resulted in the presence of XG FTase activity in these cells. AtFUT1 shows very little identity with FucTs from other organisms. AtFUT1 and PsFUT1 (the pea XyG FucT homologue) are 62.3% identical (117). Both enzymes contain motifs that had been identified in other FucTs but combine these motifs in a unique manner (117). Three motifs had been identified in (1→2) α - and (1→6) α -FucTs. Motifs I and II had been present in both (1→2) α - and (1→6) α -enzymes, but a particular version of motif III had appeared to be characteristic of each group. AtFUT1 and PsFUT1, however, contain a hybrid motif III that has features of both the (1→2) α - and (1→6) α -versions. There are ten genes in Arabidopsis with identity of encoded amino acid sequences to AtFUT1 ranging from 35 to 73.8% (118). The AtFUT1 gene was found to correspond to the fucose-deficient *mur2* mutant of Arabidopsis (119).

The galactosyltransferase that contributes to the synthesis of XG side chains was identified by map-based cloning of the *MUR3* gene of Arabidopsis, which had previously been identified based on a screen for variation in cell wall polysaccharide composition (120). *MUR3* belongs to a large family of Type II membrane proteins that is evolutionarily conserved among higher plants. The enzyme shows sequence similarities to the glucuronosyl transferase domain of exostosins, a class of animal glycosyltransferases that catalyze the synthesis of heparan sulfate, a glycosaminoglycan with numerous roles in cell differentiation and development. Arabidopsis has ten genes encoding proteins with significant sequence similarity to the *MUR3* xyloglucan GalT (121).

One of the XG xylosyltransferases (XT1) was identified from Arabidopsis based on sequence similarity to the fenugreek mannan α -1,6-galactosyltransferase (122). Expression of the gene in *Pichia pastoris* resulted in a protein with cello-oligosaccharide-dependent xylosyltransferase activity. Characterization of the products obtained with cellopentaose as acceptor indicated that the pea and the Arabidopsis enzymes transfer xylose mainly to the second glucose residue from the non-reducing end in an α (1,6)-linkage to the glucan chain. Arabidopsis has seven related genes, some of which may catalyze addition of xylose to other positions in the repeating unit of XG.

In vitro assays of the glucan synthase involved in synthesis of the XG backbone exhibit maximal activity only if both UDP-glucose and UDP-xylose are present, suggesting that the glucan synthase acts in concert with a xylosyltransferase that adds side chains (94, 123).

The glucan synthase extends existing XG by addition to the non-reducing end but cannot be primed with exogenous primers (94). Also the enzyme does not add xylose to preformed glucans.

Recently, Cocuron and coworkers (124) have obtained evidence that proteins of the cellulose synthase-like 4 (CSLC4) family catalyze synthesis of the XG backbone (99). They expressed *CSLC4* genes from Arabidopsis and tamarind along with the Arabidopsis XT1 gene in *Pichia pastoris* and observed the formation of β -1,4-linked glucan. However, they were unable to detect XG synthase activity in extracts from the cells. Mutations in the Arabidopsis *CSLC4* gene are deficient in xyloglucan, supporting the proposed role in xyloglucan synthesis (Milne and Somerville, unpublished). If substantiated by further work, the work of Cocuron and coworkers (124) appears to represent a long-awaited breakthrough in understanding the synthesis of XG (99). The identification of the genes involved in XG synthesis should pave the way for an analysis of how the amount of XG is regulated and what the consequences are to plant growth and development and the properties of cell walls of genetic variation in the amount of XG.

Some information about genetic variation in XG is available from analysis of Arabidopsis mutants that were recovered by screening for alterations in total cell wall sugar composition (125, 126). Comparison of the mechanical responses of *mur2* (AtFUT1) and *mur3* (XG galactosyltransferase), indicated that galactose-containing side chains of xyloglucan make a major contribution to overall wall strength, whereas xyloglucan fucosylation plays a comparatively minor role (127). Thus, it seems unlikely that it will be possible to develop biomass feedstock crops with significant alterations in the structure of XG without also making compensating changes in another cell wall component. Because Arabidopsis has a number of CSLC genes, it has not yet been possible to develop mutant plants with major reductions in the amount of XG to assess the phenotypic consequences of such alterations.

5.3.3 Xylan

Xylan, a polymer of β -(1-4)-linked D-xylose is one of the main components of woody plants. Xylans are usually substituted by side chains of arabinose or glucuronic acid and may be acetylated. Thus, glucuronoxylan (GX) is composed of a linear backbone of β -(1-4)-linked D-xylosyl (Xyl) residues, some of which bear a single α -D-glucuronic acid (GlcA) or 4-O-methyl- α -D-glucuronic acid (MeGlcA) residue at O-2. The Xyl residues can also be substituted with arabinosyl and acetyl residues (128). Xylosyltransferase and glucuronyltransferase activities have been detected in numerous plants (129). However, none of the genes encoding these enzymes has been identified, nor have any of the enzymes been purified to homogeneity and biochemically characterized.

Mutations in three glycosyltransferases, *FRAGILE FIBER8* (*FRA8*), *IRREGULAR XYLEM8* (*IRX8*), and *IRX9*, have been shown to be required for normal vessel morphology and wall thickness and for normal amounts of xylose and cellulose in cell walls (73, 74, 130, 131). These genes are specifically expressed in cells undergoing secondary wall thickening. Plants carrying mutations in these genes have reduced amounts of wall GX and a decreased ratio of GlcA to MeGlcA residues in the GX (129, 130, 132).

IRX8, *IRX9*, and *FRA8* are specifically expressed in fibers and vessels and their encoded proteins are localized in the Golgi (129, 130, 132). Thus, they have the properties expected of enzymes involved in glucuronoxylan synthesis. However, it has not been possible to directly

associate enzyme activity with the proteins. Thus, it is not clear how they participate in the synthesis of xylan. Peña and coworkers (129) showed that the glycosyl sequence **4-β-D-Xylp-(1 → 4)-β-D-Xylp-(1 → 3)-α-L-Rhap-(1 → 2)-α-D-GalpA-(1 → 4)-D-Xylp** was present at the reducing end of Arabidopsis GX, as previously noted for birch (*Betula verrucosa*) and spruce (*Picea abies*) wood. They further noted that mutations in *IRX8* and *IRX9*, and by inference *FRA8*, lead to reductions in the amount of the GX reducing end sequence suggesting that these genes participate in the synthesis of the GX reducing end sequence and suggest that *IRX9* has an essential role in the elongation of the xylan backbone.

The *fra8* gene encodes a GT47 family enzyme and expression of the poplar (*Populus alba* × *tremula*) *GT47C* gene in *fra8* plants rescues the defects in secondary wall thickness and GX synthesis, suggesting that *GT47C* is a functional homolog of *FRA8* (133). The *FRA8* gene encodes a putative GT in family GT47 (130). This family includes enzymes with an inverting mechanism, which usually leads to β-glycosidic linkages (when typical α-linked donor substrates are used). Thus, if UDP-α-D-Xyl is the donor substrate, it is possible that *FRA8* catalyzes the formation of the β-linkage of xylose to either O-3 of the rhamnose or O-4 of the penultimate xylose of the GX reducing end glycosyl sequence (129). However, in plants, the addition of α-Rha residues is catalyzed by inverting GTs that use UDP-β-L-Rha as the donor substrate (134). Therefore, it is also possible that *FRA8* catalyzes the addition of rhamnose during the biosynthesis of the GX reducing end sequence.

The *IRX8* (*GAUT12*) gene encodes a putative GT in family GT8 (73, 74, 129, 132). Several members of the GT8 family catalyze the transfer of uronic acids to glycans. For example, three Arabidopsis GT8 proteins, QUASIMODO1 (*QUA1*) (135), PARVUS (136), and GALACTURONOSYLTRANSFERASE1 (*GAUT1*) (137), have been identified and are believed to have a role in pectin biosynthesis. Of these three, only *GAUT1* has been biochemically characterized and shown to have galacturonosyltransferase activity (137). Family GT8 enzymes are retaining glycosyl transferases that catalyze the formation of α-glycosidic bonds when using α-linked donor substrates such as UDP-α-D-GalA. Thus, it is possible that *IRX8* catalyzes the addition of an α-D-GalA residue to O-4 of the reducing Xyl residue present in the GX reducing end sequence described above.

Mutation of the *IRX9* gene, which encodes a putative GT in family GT43 (138), was shown to result in plants with decreased amounts of wall GX, suggesting that this gene is required for GX synthesis (131). The poplar (*Populus tremula* × *tremuloides*) *GT43A* and Ptt *GT43B* genes, which are homologs of *IRX9*, have been shown to be highly expressed during wood formation (139). In addition, a cotton (*Gossypium hirsutum*) gene, which resides in the same phylogenetic subgroup as Ptt *GT43A*, Ptt *GT43B*, and *IRX9*, is highly expressed during cotton fiber development (140). Together, these findings suggest that family 43 GTs have an important role in secondary wall synthesis. Enzymes in this family are distinguished by an inverting mechanism, typically catalyzing the formation of β-glycosidic bonds using α-linked glycosyl donors. Our demonstration that the *irx9* mutation leads to a decrease in the chain length of GX suggests that *IRX9* encodes a xylan synthase responsible for adding β-xylosyl residues to the nascent GX. This hypothesis is consistent with our results indicating that *IRX9* is highly expressed in cells undergoing secondary wall biogenesis and that *IRX9* is localized in the Golgi, where GX synthesis occurs (141, 142). However, it is also possible, as we discussed previously (130), that an inverting GT can catalyze the formation of an α-linkage when a β-linked substrate (such as a glycosyl phospholipid) is used as the donor. Such inverting enzymes can also catalyze the formation of high-energy β-linked glycosides (such as glycosyl phospholipids) that are subsequently used as glycosyl donors. Thus, an

alternative interpretation is that IRX9 is directly or indirectly involved in the transfer of α -linked GlcA residues to the GX backbone. Additional studies are required to determine whether IRX9 catalyzes the addition of xylose or GlcA to the GX backbone.

In the lignified walls of the Poaceae, the major non-cellulosic polysaccharides are glucuronoarabinoxylans (GAXs), although the degree of substitution of the xylan main chain is less than in the GAXs of the primary cell walls. In the non-lignified walls of the Poaceae and other species such as pineapple (*Ananas comosus*), ferulic acid is ester-linked to GAXs. These polysaccharides comprise only a minor component of the non-cellulosic polysaccharides of the non-lignified walls of species in the basal Arecales (palms) clade (143), but are a major component of the non-lignified walls of species in the other commelinid clades, particularly the Poales (144–146).

5.3.4 Mixed linkage glucans

Walls of the grasses contain mixed-linkage (1,3;1,4)- β -D-glucans (MLGs) which are not present in walls of dicotyledons or most other monocotyledonous plants (147). Some alga and liverworts may also have MLGs (148). The (1,3;1,4)- β -D-glucans have an unusual structure consisting of an unbranched, unsubstituted glucan chain with two linkages arranged in a non-repeating, but non-random, fashion. The glucan chains consist of primarily celotriosyl and cellotetraosyl units separated by single (1 \rightarrow 3)- β -linkages (149). MLGs can be synthesized in vitro from Golgi membrane fractions with UDP-Glc as a substrate (150, 151). The amount of UDP-Glc used in the assay alters the ratio of celotriosyl and cellotetraosyl units, indicating that this is not a fixed property of the biosynthetic enzyme (152).

Recently, following prescient speculation as to possible structural similarity between cellulose synthase and the MLG synthase (153), it was shown that expression of a rice *CsLF* gene in *Arabidopsis* led to the accumulation of (1,3;1,4)- β -D-glucan biosynthesis in *Arabidopsis* (154). Thus, it appears that CSLF encodes the mixed glucan synthase. The generally low levels of (1,3;1,4)- β -D-glucan in walls of the transformed *Arabidopsis* plants is consistent with the concept that limiting levels of other components might be required for high-level synthesis of the polysaccharide or its transfer to the cell wall. Similarly, the preferential deposition of the (1,3;1,4)- β -D-glucan in the epidermal layers of the transgenic *Arabidopsis* lines, despite the fact that transgene expression was driven by the constitutive 35S promoter, could indicate that the epidermal cells contain ancillary factors that are not abundant in other cells of the leaf. The identification of the gene opens up new approaches to understanding the fascinating process by which an enzyme catalyzes substantially different transferase reactions (151, 153).

The role of the MLGs is not clear. The MLGs are synthesized in relatively large amounts during growth and may coat cellulose microfibrils during the synthesis and expansion phase, but they are degraded when elongation ceases (155, 156).

5.4 Pectins

Pectin is likely the most structurally complex family of polysaccharides in nature (Figures 5.2 and 5.3). Pectin is particularly abundant in primary walls, i.e. those walls surrounding

Homogalacturonan

4) α GalA-(1,4)- α GalA-(1,4)- α GalA-(1,4)- α GalA-(1,4)- α GalA-(1,4)- α GalA-(1,4)- α GalA-(1,

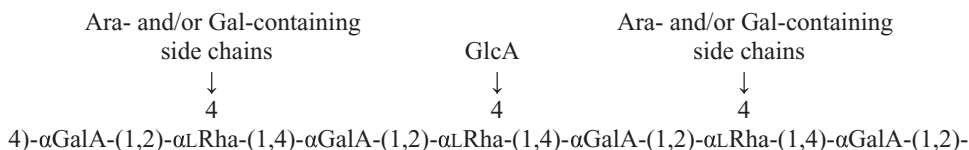
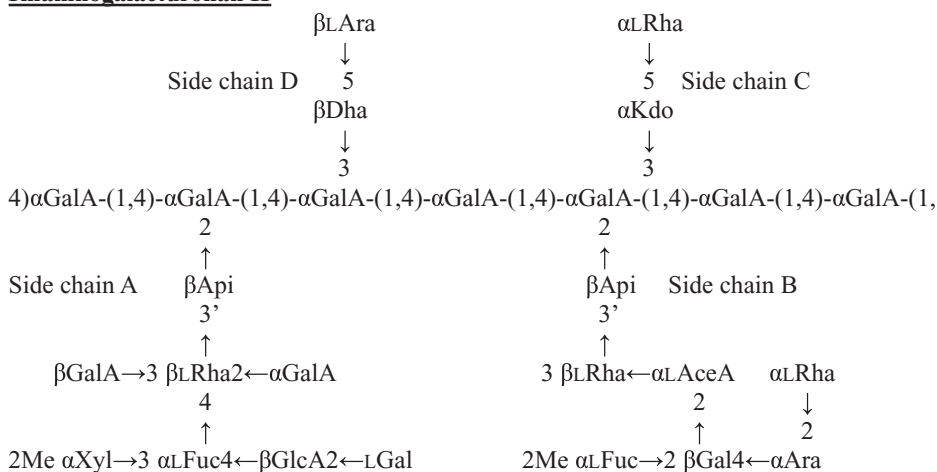
Rhamnogalacturonan I**Rhamnogalacturonan II**

Figure 5.2 Representative structures of the three pectic polysaccharides HG, RG-I, and RG-II.

growing and dividing cells and the terminal wall in many cells of the soft parts of the plant. Pectin is also abundant in the middle lamella which is the junction between adjacent cells. Pectin comprises ~35% of the polysaccharides in dicot and non-graminaceous monocot primary walls, and 2–10% of the wall in the grasses (157, 158). Pectin is also present in the walls of gymnosperms, pteridophytes, and bryophytes as well as Chara, a charophycean alga, which is believed to be the closest extant relative of land plants (159). Although pectin is not a major component of secondary walls, it is present as the outer layer of secondary walls and can represent ~5% of harvested tree wood. Thus, depending on the plant and tissue used, pectin will be present in the biomass used for biofuel production and, since it comprises a complex interconnected matrix in the wall, likely affects the recalcitrance of biomass to deconstruction for biofuel production.

Pectins have multiple roles in plant defense, growth, and development (158, 160). They provide wall structure (161), bind and exchange apoplastic anions and macromolecules (162, 163), influence cell–cell adhesion (164, 165), and are involved in cell signaling (166,

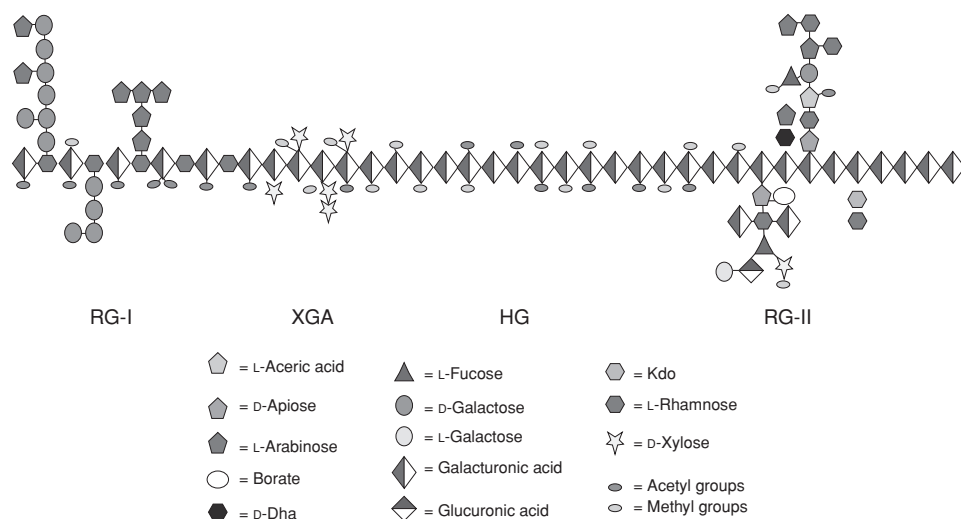


Figure 5.3 Schematic structure of pectin showing the three main pectic polysaccharides homogalacturonan (HG), rhamnogalacturonan I (RG-I), and rhamnogalacturonan II (RG-II) linked to each other. A region of substituted galacturonan known as xylogalacturonan is also shown (XGA). The representative pectin structure shown is not quantitatively accurate, HG should be increased 12.5-fold and RG-I increased 2.5-fold to approximate the amounts of these polysaccharides relative to each other plant walls. The monosaccharide symbols used are either from the Symbol and Text Nomenclature for Representation of Glycan Structure. Nomenclature Committee Consortium for Functional Glycomics (<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>) or from D. Mohnen. (The figure is modified from <http://www.uk.plbio.kvl.dk/plbio/cellwall.htm>.) (Reproduced in color as Plate 4.)

167). Pectins have roles in pollen tube growth (168), seed hydration (169–171), leaf abscission (172), guard cell function (173), organ formation (174, 175), fruit development (158), and possibly water movement (176). Pectic oligosaccharides are intercellular signal molecules (177) in plant development (166) and defense responses (178, 179). Mutant plants with altered pectin structure may be dwarfed (161, 180), have brittle leaves (164), reduced numbers of side shoots and flowers (175), and reduced cell–cell adhesion (135, 181).

Pectin is defined as a family of plant cell wall polysaccharides that contain 1,4-linked galacturonic acid (157). Galacturonic acid (GalA) comprises roughly 70% of total cell wall pectin and is a major component of the three major types of pectic polysaccharides: homogalacturonan (HG), rhamnogalacturonan I (RG-I), and the substituted galacturonans for which rhamnogalacturonan II (RG-II) is the most ubiquitous and structurally invariant member (157). In addition, pectin includes the less abundant substituted galacturonan xylogalacturonan (XGA) (182–187) and apiogalacturonan (AG) (158, 188–191). The complex structure of the pectic polysaccharides makes the study of pectin synthesis challenging. It is estimated that at least 58 enzymes are required to synthesize pectins, including methyltransferases, acetyltransferases and numerous glycosyltransferases (192).

HG accounts for ~65% of pectin (193, 194) and is a homopolymer of α -D-1,4-linked GalA residues (Figures 5.2 and 5.3) that is partially methylesterified at the C-6 carboxyl (157, 195), may be O-acetylated at O-2 or O-3 (196–199), and may contain other esters

whose structure remains unclear (200–204). RG-I accounts for 20–35% of pectin (194) and is a family of polysaccharides with an alternating $[\rightarrow 4)\text{-}\alpha\text{-D-GalA-(1}\rightarrow 2)\text{-}\alpha\text{-L-Rha-(1}\rightarrow]$ backbone (Figures 5.2 and 5.3). Between 20 and 80% of the rhamnosyl residues are substituted with side chains composed predominantly of linear and branched $\alpha\text{-L-Araf}$ and $\beta\text{-D-Galp}$ (157, 205). The main types of side chains include $\alpha\text{-1,5-linked L-arabinan}$ with some 2- and 3-linked arabinose or arabinan branching, $\beta\text{-1,4-linked D-galactans}$ with some 3-linked L-arabinose or arabinan branching and $\beta\text{-1,3-linked D-galactan}$ with $\beta\text{-6-linked galactan}$ or arabinogalactan branching (205). RG-I side branches may also contain $\alpha\text{-L-Fucp}$, $\beta\text{-D-GlcpA}$, and 4- *O*-Me $\beta\text{-D-GlcpA}$ residues (206). The composition and length of RG-I side chains varies between cell types and in different plant species (158, 160). RG-II accounts for $\sim 10\%$ of pectin (158, 194) and contains 12 different types of sugars in over 20 different linkages. The HG backbone of RG-II is substituted at O-2 and O-3 with four structurally complex oligosaccharides A–D (159) (Figures 5.2 and 5.3). RG-II in the plant generally occurs as a RG-II dimers crosslinked by borate diesters (159). The 4-linked galacturonans that are substituted at O-3 with D-xylose (the xylogalacturonans, XGA) are often found in reproductive tissues (157, 186, 193, 207) whereas galacturonans substituted at O-2 or O-3 by D-apiofuranose (188, 189) (the apiogalacturonans, AG) are restricted to selected aquatic monocots (e.g., Lemna).

When walls are isolated from the plant, the pectic polysaccharides appear to be covalently cross-linked since harsh chemical treatments or digestion by pectin-degrading enzymes is required to isolate HG, RG-I, and RG-II separately from each other. It is not known, however, how the pectic polysaccharides are covalently linked to each other or to other polymers in the wall. It is also not clear where and how that crosslinking occurs, i.e. via the action of glycosyltransferases in the Golgi or by transglycosylases or other enzymes in the wall. The available data (208) support a model whereby HG, RG-I, and RG-II are linked via their backbones. However, due in part to the uncertainty of how and where the pectic polysaccharides are cross-linked, it is currently not possible to predict the complete repertoire of biosynthetic enzymes that are needed to synthesize pectin. Furthermore, although the general types of pectic polysaccharides are similar in different plant species, there is a growing body of evidence showing that species-, cell-type-, and developmental state-specific differences in pectin structure exist, thereby making it likely that the number and types of enzymes required to synthesize pectin will depend on the plant, tissue and developmental state of the cells of interest.

Finally, a knowledge of the structure of the “mature” polysaccharides in the wall, or at least those that can be isolated from the wall and characterized, does not necessarily reflect the structures as they are synthesized, but rather the structures as they are inserted into the wall and after they have been modified by wall-localized enzyme catalyzed (and chemical) reactions. In the following discussion of pectin, there is no attempt to define species-specific differences in pectin synthesis, since our understanding of the species-specific tailoring of pectin structures and synthesis is only just beginning to be studied. Rather, this review emphasizes our current understanding of biosynthetic enzymes required for the basic pectin structures that appear to be common in all species. Although only few of the genes encoding pectin biosynthetic enzymes have been confirmed by demonstration of enzymatic activity of the encoded proteins, recent progress in identifying genes encoding putative pectin biosynthetic enzymes make it likely that more pectin biosynthetic genes will be functionally identified in the near future. The availability of such genes should facilitate the elucidation

of how diverse pectin biosynthetic enzymes work together, likely within protein complexes (Atmodjo and Mohnen, unpublished results) to synthesize the multifunctional family of pectic polysaccharides.

Several comprehensive reviews on pectin biosynthesis (158, 192, 205, 209), as well as more general reviews on plant wall biosynthesis (72, 99, 126, 155, 210–213, 214) and strategies to identify wall biosynthetic glycosyltransferases (118, 215–218) and regulation of cell wall synthesis (219–221) have previously been published. This review will attempt to merge recent advances in pectin synthesis with the prior studies so as to reflect our current understanding of pectin synthesis.

5.4.1 Location of pectin synthesis

All available evidence, including autoradiographic pulse chase studies using wall biosynthetic precursors (222, 223), immunocytochemical studies using anti-pectin-specific antibodies (224–226), and subcellular fractionation and topology studies of pectin biosynthetic enzymes (227–231), indicate that pectin is synthesized in the Golgi and transported to the wall in membrane vesicles. Plant cells, unlike animal cells, have multiple Golgi and thus pectin synthesis occurs simultaneously in numerous Golgi stacks in the cell (225, 232). The synthesized pectin and other macromolecules are targeted to the wall by the movement of Golgi vesicles, presumably along actin filaments that have myosin motors (233).

Immunocytochemical studies also indicate that the synthesis of different regions of the pectic polysaccharides occurs in different Golgi cisternae as pectin moves from the cis, through the medial and to the trans-Golgi. For example, the use of antibodies specific to different regions of HG and RG-I suggests that HG and RG-I synthesis begins in the cis-Golgi (225, 234, 235) and continues with more extensive decoration of the backbones as the polymers move through the medial Golgi (224, 225, 235) and into the trans-Golgi cisternae (225, 235). Additional modifications of the pectic glycan structure also appear to proceed in a more-or-less organized manner with HG (236, 237) and RG-I (106, 234, 238, 239) initially synthesized in less modified forms in the cis- and medial-Golgi and becoming more modified (e.g., methylesterified) (236) in the medial- and trans-Golgi (225, 235, 240–242). HG is believed to be transported to the plasma membrane and inserted into the wall as a highly methylesterified polymer (214, 237, 243–245) and once in the wall, HG is deesterified to varying degrees by pectin methylesterases (246) in the wall or at the cell plate (245). The deesterification of HG converts it to a more negatively charged form (240, 247–250) which is then available to bind ions, enzymes, proteins, and other HG molecules through Ca^{++} salt bridges. It is believed that a spatial partitioning of HG esterification and deesterification occurs in the wall based on localization of esterified HG throughout the cell wall (237, 240–242, 243, 245, 250, 251), while relatively unesterified HG is more restricted to the middle lamella. This conclusion is supported by the frequently observed absence of unesterified HG epitopes in the trans-Golgi vesicles. However, since some cell types, such as melon callus cells (240), contain unesterified HG in the trans-Golgi, it is possible that HG may be inserted into the wall in a relatively unesterified form, at least in some cells. Also, since specific pectic epitopes localize to different Golgi compartments in different cell types (5, 225, 234, 237, 244), it is likely that the specific localization of the diverse pectin biosynthetic enzymes may vary in a cell type, species, and development-specific manner (226, 252–255). It must be noted, however, that the interpretation of immunocytochemistry

results can be difficult since the absence of a signal using an epitope-specific antibody may be due to masking of the epitope by additional glycosylation or some other modification (e.g., methylation, acetylation, feruloylation). Thus, to conclude that a particular pectin biosynthetic event does not occur in a cell, the lack of a particular immunocytochemical signal is not sufficient. Information on the presence of the biosynthetic enzyme activity or the actual wall carbohydrate structure itself is required.

5.4.2 *Pectin biosynthetic glycosyltransferases*

The current view is that the bulk of pectin synthesis is catalyzed by glycosyltransferases (GTs) that transfer a glycosyl residue from an activated form of the sugar, most likely a nucleotide-sugar, to an acceptor. Like other polymer biosynthetic reactions, pectin synthesis is thought to proceed through three stages: initiation, elongation, and termination. There is no detailed understanding of the initiation phase for the synthesis of any of the pectic polysaccharides. It has traditionally been held that the pectic polysaccharides are not synthesized on a protein, thus distinguishing them from the synthesis of animal Golgi-localized proteoglycans (256). More recently, several investigators have presented evidence suggesting that some wall polysaccharide synthesis may occur on protein primers (257–259); however, definitive proof of this hypothesis has yet to be provided. To date all studies of pectin biosynthetic glycosyltransferases have been carried out by assaying the transfer of a glycosyl residue from a radiolabeled nucleotide-sugar substrate (with or without additional unlabeled substrate) onto endogenous acceptors in plant microsomal membranes or onto exogenous acceptors, or alternatively, by transfer of an unlabeled substrate onto fluorescently labeled exogenously added oligosaccharide or polysaccharide acceptors. Most nucleotide-sugars involved in pectin synthesis consist of a nucleoside-diphosphate (NDP) attached to the sugar, and thus the general reaction catalyzed by the GTs is $\text{NDP-sugar} + \text{acceptor}_{(n)} \rightarrow \text{NDP} + \text{acceptor}_{(n+1)}$. The precise number and location of the glycosyl residues in the acceptor that are recognized by a GT will be unique for each GT. However, in this review, for the purposes of calculating a minimal number of GTs required for pectin synthesis, since all pectin biosynthetic GTs characterized to date have been shown to add onto the non-reducing end of the oligosaccharide/polysaccharide acceptor, the assumption has been made that each GT will recognize, as a minimum, the terminal two glycosyl residues at the non-reducing end of the acceptor (i.e., the sugar onto which the transfer takes place, and the adjacent sugar).

Table 5.2 lists the types of glycosyltransferases that are expected to be required for pectin synthesis, based on the premise that a unique glycosyltransferase will be required for the transfer of a unique sugar from a unique nucleotide-sugar onto a unique disaccharide acceptor region at the non-reducing end of the acceptor. In some cases, it is possible that the same enzyme may catalyze the synthesis of a similar region on diverse polysaccharides, (e.g., the same galacturonosyltransferase may catalyze the synthesis of the backbone of HG and the HG region of RG-II. Table 5.2 attempts to list all known or expected GTs that are required for pectin synthesis, and as such, is meant to serve as a reference table. However, in an effort to consider in more depth the synthesis of the different types of pectic polysaccharides, HG, RG-I, RG-II, XGA, and AG; a detailed summary of progress in understanding the synthesis of the specific pectic polysaccharides is described separately.

Table 5.2 List of glycosyltransferases expected to be required for pectin biosynthesis

Type of glycosyl-transferase	Parent polymer ^a (side chain)	Enzyme ^b acceptor substrate activity (unless noted: enzyme adds to the glycosyl residue on the left*)	Ref. ^c for structure	Putative gene Identified (Ref.) ^d	Gene identified (Ref.) ^e
d-GalAT	HG/RG-II	*GalAα1,4-GalA α1,4-GalAT	(157)	Put. QUA1- AT3g25140 (135, 260)	GAUT1 AT3g61130 (137)
d-GalAT	RG-I	L-Rhaα1,4-GalA α1,2-GalAT	(157, 261, 262)	—	—
d-GalAT	RG-II (A)	L-Rhaβ1,3'-Apiif α1,2-GalAT	(214, 157)	—	—
d-GalAT	RG-II (A)	L-Rhaβ1,3'-Apiif β1,3GalAT	(214, 157)	—	—
d-GalAT	RG-I/HG	GalAα1,2-L-Rha α1,4-GalAT	—	—	—
d-GalAT	HG attached to xylan	*GalAα1,4-GalA α1,4-GalAT or Xyl-?-.... α1,4-GalAT (Xylβ1,4-Xylα1, 3-Rhaα1,2-GalA-1,4-Xyl)	(129, 263)	Put. GAUT12 IRX8 AT5g54690 (GT8) (129, 132)	—
L-RhaT	RG-I	GalAα1,2-L-Rha α1,4L-RhaT	(157, 261, 262)	—	—
L-RhaT	RG-II (A)	Apiifβ1,2-GalA β1,3'-L-RhaT	(214, 157)	—	—
L-RhaT	RG-II ©	Kdo2,3GalA α1,5-L-RhaT	(214, 157)	—	—
L-RhaT	RG-II (B)	L-Araα1,4-Gal α1,2-L-RhaT	(214, 157)	—	—
L-RhaT	RG-II (B) ^f	L-Araα1,4-Gal β1,3-L-RhaT	(158, 161)	—	—
L-RhaT	HG/RG-I	GalAα1,4-GalA α1,4-L-RhaT	—	—	—
d-GalT	RG-I	L-Rhaα1,4-GalA β1,4-GalT	(157, 261)	—	—
d-GalT	RG-I	Galβ1,4-Rha β1,4-GalT	(157, 264)	—	—
d-GalT	RG-I	Galβ1,4-Gal β1,4-GalT	(157, 190, 207, 264-267)	—	—
d-GalT	RG-I	Galβ1,4-Gal β1,6-GalT	(157, 264)	—	—
d-GalT	RG-I/AGP ^g	Galβ1,3-Gal β1,3-GalT	(214)	—	—
d-GalT	RG-I/AGP	Galβ1,3-Gal β1,6-GalT	(214)	—	—
d-GalT	RG-I/AGP	Galβ1,6-Gal β1,3-Gal β1,6-GalT	(214)	—	—
d-GalT	RG-I	L-Araf1,4-Gal 1-5-GalT	(268)	—	—
L-GalT	RG-II (A)	GlcAβ1,4-Fuc α1,2-L-GalT	(214, 269)	—	—
d-GalT	RG-II (B)	L-Acef A α1,3-Rha β1,2GalT	(214, 269)	—	—
L-AraT	RG-I	Galβ1,4-Rha α1,3-L-AraT	(157, 264)	—	—

L-AraT	RG-I	L-Araf α 1,3-Gal	α 1,2-L-ArafT	(157, 264)	—	—
L-AraT	RG-I	L-Araf α 1,2-Araf	1,5-L-ArafT	(157, 264)	—	—
L-AraT	RG-I	L-Rha	α 1,4-GalA	(264)	—	—
L-AraT	RG-I	L-Araf α 1,5-Araf	α 1,5-L-ArafT	(214)	Put. ARAD1	—
L-AraT	RG-I	L-Araf α 1,5-Araf	α 1,2-L-ArafT	(214)	At2g35100 (270)	—
L-AraT	RG-I	L-Araf α 1,5-Araf	α 1,3-L-ArafT	(214)	—	—
L-AraT	RG-I	L-Araf α 1,3-Araf	α 1,3-L-ArafT	(214)	—	—
L-AraT	RG-I	Gal β 1,4-Gal	α 1,3-L-ArafT	(157, 190, 207, 214, 265, 266)	—	—
L-AraT	RG-I	L-Araf1,3-Gal	1,5-L-ArafT	(190, 207, 265, 266)	—	—
L-AraT	RG-I/AGP	Gal β 1,6-Gal	α 1,3-L-ArafT	(214)	—	—
L-AraT	RG-I/AGP	Gal β 1,6-Gal	α 1,6-L-ArafT	(214)	—	—
L-ArapT	RG-I/AGP	L-Araf α 1,5-Araf	β 1,3-L-ArapT	(271, 272)	—	—
L-ArapT	RG-I	Gal β 1,4-Gal	α 1,4-L-ArapT	(273)	—	—
L-AraT	RG-II (D)	Dha β 2,3-GalA	β 1,5-L-ArafT	(214, 269)	—	—
L-AraT	RG-II (B)	Gal β 1,2-L-AcefA	α 1,4-L-ArapT	(214, 269)	—	—
L-AraT	RG-II	L-Rha α 1,2-L-Ara	β 1,2-L-ArafT	(214, 269)	—	—
L-FucT	RG-I	Gal β 1,4-Gal	α 1,2-L-FucT	(157, 264)	—	—
L-FucT	RG-II (A)	L-Rha β 1,3'-Apif	α 1,4-L-FucT	(214, 269)	—	—
L-FucT	RG-II (B)	Gal β 1,2-L-AceAf	α 1,2-L-FucT	(214, 269)	—	—
D-ApifT	RG-II (A, B)	GalA α 1,4-GalA	β 1,2-ApifT	(214, 269)	—	—
D-XylT	RG-II (A)	L-Fuc α 1,4-L-Rha	α 1,3-XylT	(214, 269)	—	RGXT1
					At4g01770,	
					RGXT2	
					At4g01750,	
					(274)	
D-XylT	HG	GalA α 1,4-GalA	β 1,3 XylT	(185, 186, 193, 207, 263)	—	—
D-XylT	HG	Xyl β 1,3-GalA	β 1,4 XylT	(263)	—	—
D-XylT	HG	Xyl β 1,4-Xyl β	1,3-GalA β 1,4 XylT	(263)	—	—

(continued)

Table 5.2 (continued)

Type of glycosyl-transferase	Parent polymer ^a (side chain)	Enzyme ^b acceptor substrate activity (unless noted: enzyme adds to the glycosyl residue on the left*)	Ref. ^c for structure	Putative gene Identified (Ref.) ^d	Gene identified (Ref.) ^e
d-XylIT	HG	Xyl[β1,4-Xyl] β1,4 XylIT	(263)	–	–
d-XylIT	RG-I	L-Rhaα1,4-GalA β1,4 XylIT	(275–277)	–	–
d-XylIT	RG-I	Xyl[β1,4-L-Rhaα1,4-GalA β1,4 XylIT	(277)		
d-GlcAT	RG-I	Gal..... β1,6GlcAT	(278)	Put. <i>NpGUT1</i> (181)	
d-GlcAT	RG-I	Gal..... β1,4GlcAT	(278)	–	–
d-GlcAT	RG-II (A)	L-Fucα1,4-L-Rha β1,4GlcAT	(214, 269)	–	–
d-KDOT	RG-II (C)	Gal/Aα1,4-GalA α2,3KdoT	(159, 214, 269)	–	–
d-DHAT	RG-II (D)	Gal/Aα1,4-GalA β2,3DhaT	(214, 269)	–	–
L-AcefA	RG-II (B)	L-Rhaβ1,3'-Apif αL,3AcefAT	(214, 269)	–	–

^a HG, homogalacturonan; RG-I, rhamnogalacturonan I; RG-II, rhamnogalacturonan II.
^b All sugars are D sugars and have pyranose rings unless otherwise indicated.
^c References for the structure.
^d Put: putative, indicates that a gene encoding the putative corresponding GT has been identified, but confirmatory functional enzyme activity of the gene has not yet been provided.
^e References for gene encoding the GT when available.
^f Glycosyltransferase activity based on the most extended structure of RG-II. The terminal βRhap → 3αArap in side-chain B and the terminal βAraf → 2αRhap in side-chain B are not present in RG-II from all species (159).
^g Enzyme activity would also be required to synthesize arabinogalactan proteins (AGPs) [see (279)].

Table 5.3 List of non-glycosyltransferases expected to be required for pectin synthesis

Type of transferase	Parent polymer ^a	Enzyme activity	Enzyme acceptor ^b substrate	Ref. ^c	Putative gene identified (Ref.) ^d
MethylIT	HG	HG-methyltransferase	GalA α 1,4-GalA _(n)	(227, 281, 286, 287)	Put. At1g78240 (QUA2) (287)
AcetylIT	HG	HG: GalA 3- <i>O</i> -acetyltransferase	GalA α 1,4-GalA _(n)	(196–199)	
AcetylIT	RG-I	RG-I: GalA-3- <i>O</i> /2- <i>O</i> -acetyltransferase	GalA α 1,2-L-Rha α 1,4 _(n)	(157, 199, 288–290)	
MethylIT	RG-I	RG-I: GlcA-4- <i>O</i> -methyltransferase	GlcA β 1,6-Gal	(278)	
MethylIT	RG-II	RG-II: xylose-2- <i>O</i> -methyltransferase	D-Xyl α 1,3-L-Fuc	(214, 269)	
MethylIT	RG-II	RG-II: fucose-2- <i>O</i> -methyltransferase	L-Fuc α 1,2-D-Gal	(214, 269)	
AcetylIT	RG-II	RG-II: fucose-acetyltransferase	L-Fuc α 1,2-D-Gal	(214, 269)	
AcetylIT	RG-II	RG-II: aceric acid 3- <i>O</i> -acetyltransferase	L-AcefA β 1,3-L-Rha	(214, 269)	

^a HG, homogalacturonan; RG-I, rhamnogalacturonan I; RG-II, rhamnogalacturonan II.

^b All sugars are D sugars and have pyranose rings unless otherwise indicated.

^c Reference is for the enzyme activity, when available.

^d Put.: putative, indicates that a possible gene encoding the corresponding GT has been identified, but confirmatory functional enzyme activity of the gene has not yet been provided.

5.4.3 Methyltransferases

Many of the pectin polysaccharides contain sugars that have been modified by the addition of a methyl group. For example, HG has carboxymethyl esters on C-6, RG-I contains a GlcA-4-*O*-methylether, and RG-II contains 2-*O*-methylfucose. Multiple methyltransferases are required to synthesize these polysaccharides, although it is not known if the methyl group is added after the sugar is inserted into the polymer, or rather on the nucleotide-sugar donor. The available evidence suggests that SAM is the methyl donor (227, 280–285). Table 5.3 lists the non-glycosyltransferases “required” for the synthesis of pectin.

5.4.4 Acetyltransferases

Several of the pectic polysaccharides contain sugars modified by acetylation. These include acetylation of the O-3 position of GalA in HG and either O-2 or O-3 acetylation on the GalA in the RG-I backbone. In addition, RG-II contains acetylated Fuc and 3-*O* acetylated aceric acid. The acetyltransferases have not been identified.

5.4.5 Other pectin modifying enzymes

Some plant cell wall polysaccharides contain small amounts of ester-linked hydroxycinnamic acid derivatives such as *p*-coumaric and ferulic acid (291). These ester-linked acid

derivatives can undergo oxidative coupling to form dehydrodimers that may lead to cell wall polysaccharide cross-linking (291, 292) and may also be involved in the formation of polysaccharide–lignin complexes (293, 294). The ester-linked hydroxycinnamic acids are more abundant in the walls of the monocotyledonous group known as the commelinids, including the grass family (295) and such walls are particularly rich in hydroxycinnamic acids linked to the hemicellulose arabinoxylan (291), and to a lesser extent the hemicellulose xyloglucan. However, ester-linked hydroxycinnamic acids have also been shown to be linked to pectins in plants such as spinach and sugar beet (291, 296–298). The types of linkages associated with pectin include the following feruloylated arabinan and (1→4)-linked D-galactosyl oligosaccharides that presumably originate from side chains from RG-I: O-(6-*O-trans*-feruloyl)-β-D-galactopyranosyl-(1→4)-D-galactose, O-(2-*O-trans*-feruloyl)-α-L-arabinofuranosyl-(1→5)-L-arabinose, and O-α-L-arabinofuranosyl-(1→3)-O-(2-*O-trans*-feruloyl-α-L-arabinofuranosyl)-(1→5)-L-arabinose [see (291) and references therein] and O-[5-*O*-(feruloyl)-Ara]-(1→5)-[2-*O*-(feruloyl)-Ara]-(1→5)-Ara (299).

The mode of synthesis of feruloylated hemicellulose and pectin has received some, albeit, limited study. There is evidence that hemicellulosic arabinoxylan can be feruloylated by both feruloyl-glucose and feruloyl-CoA precursors/substrates, although the precise role of these substrates in the synthesis of feruloylated arabinoxylan within the cell (likely in the Golgi) or in the cell wall remains unclear (300, 301). A Golgi or sub-Golgi fraction from parsley suspension-cultured cells was able to transfer ferulic acid from feruloyl-CoA onto endogenous polysaccharide acceptors. However, since the identity of the polysaccharide(s) that was feruloylated was not determined (302), it is not clear whether the enzyme activity identified was involved in pectin or hemicellulose feruloylation.

Recently, Mitchell and coworkers (303), using a bioinformatics approach to identify genes highly expressed in cereals during the late stages of arabinoxylan synthesis, identified cereal Pfam family PF02458 genes, members of the CoA-acyl transferase superfamily, as candidate feruloyltransferases. However, enzymatic confirmation that these genes are actually involved in feruloylation has not been presented.

5.4.6 Homogalacturonan synthesis

Homogalacturonan (HG) is the most abundant pectic polysaccharide. It is a homopolymer of 1,4-linked α-D-galactosyluronic acid that is partially methylesterified and may be O-acetylated at O-2 or O-3. As mentioned above, the distribution of methylesters in HG during synthesis or in the wall is not known (157), although there is evidence that the distribution of non-esterified galacturonosyl residues in the wall is not random (195). The degree of polymerization (DP) of HG, as well as the question of whether it is linear, or may be branched or cross-linked (204) remains a matter of debate. However, the DP has been estimated to range from 72 to 100 (304) or more (214). The synthesis of HG requires at least one HG:α1,4-galacturonosyltransferase (HG-GalAT) (Table 5.2), at least one HG-methyltransferase (HG-MT) (also referred to as pectin methyltransferase), and at least one HG:O-acetyltransferase (HG-AT).

5.4.6.1 HG:galacturonosyltransferase (HG:GalAT)

The only HG biosynthetic glycosyltransferase for which enzymatic function of the encoded gene has been established is the HG:GalAT called GAUT1 (Galacturonosyltransferase 1;

At3g61130) (137). The identification of GAUT1 followed an extensive study of HG: α 1,4GalAT activity in multiple plant species including in mung bean (305–308), tomato (306), turnip (306), sycamore (309), tobacco suspension (310–312), radish roots (205), pea (230, 313), Azuki bean (*Vigna angularis*) (314), petunia (315), and *Arabidopsis* (137, 316) (see Table 5.4). This GalAT activity was shown to be particulate (i.e., membrane bound) in all species studied and was measured as the transfer of GalA from UDP-GalA onto endogenous acceptors. The GalAT activity in pea was localized to the Golgi (230) with its catalytic site facing the luminal side of the Golgi (230), providing the first direct enzymatic evidence that the synthesis of HG occurs in the Golgi.

HG:GalAT activity in microsomal membranes, measured using UDP-[14 C]GalA (317, 318), incorporates the GalA moiety onto endogenous acceptors to yield relatively large molecular mass labeled products of ~ 105 kDa in tobacco microsomal membranes (310) and ≥ 500 kDa in pea Golgi (230). Cleavage of the radiolabeled product into GalA, digalacturonic acid (diGalA), and trigalacturonic acid (triGalA) by a purified endopolygalacturonase demonstrated that the product was HG (310). In tobacco, the product produced in vitro in microsomes was $\sim 50\%$ esterified (310) while the product produced in pea Golgi was less esterified (230), suggesting that the degree of methyl esterification of newly synthesized HG may be species-specific and that methylesterification may occur after the synthesis of at least a short stretch of HG. GalAT activity can also be studied in detergent-permeabilized microsomes to allow access of the enzymes to exogenous pectic acceptors. For example, detergent-permeabilized microsomes from etiolated azuki bean seedlings transfer [14 C]GalA from UDP-[14 C]GalA onto acid-soluble polygalacturonate (PGA) exogenous acceptors (314). The azuki bean enzyme exhibited a broad pH range of 6.8–7.8 and a surprisingly high-specific activity of $1300\text{--}2000$ pmol mg^{-1} min^{-1} , considering the large amount ($3.1\text{--}4.1$ nmol mg^{-1} min^{-1}) of polygalacturonase activity that was also present in the microsomal preparations.

Success in identifying the gene encoding a GalAT required solubilizing GalAT activity from membrane preparations so as to facilitate purification of the enzyme. The first solubilization of an HG:GalAT was achieved with tobacco GalAT (311). Detergent-solubilized GalAT adds GalA onto the non-reducing end (312) of exogenous HG with a preference for HG oligosaccharides (oligogalacturonides; OGAs) of a DP of greater than 9 (311, 315), although OGA acceptors as small as a trimer can be used (315, 319). Although detergent-solubilized GalAT can use polymeric pectin substrates such as polygalacturonic acid and pectin, such polymers are less favorable substrates (314).

Studies carried out under conditions that provide information regarding the mode of elongation of the OGAs by GalAT, i.e., with excess OGA acceptor to UDP-GalA ratios (320), suggest that solubilized tobacco, radish and *Arabidopsis* enzymes, and permeabilized pea Golgi galacturonosyltransferase, have a distributive (non-processive) mode of action in vitro (230, 311, 321). Under these conditions, the bulk of the HGA elongated in vitro by solubilized GalAT from tobacco membranes (311), or detergent-permeabilized Golgi from pea (230), is elongated by a single GalA residue. As expected, as the UDP-GalA:OGA ratio is increased, the OGA products become progressively longer (137), but it is important to note that this, in itself, does not denote processivity, it simply means that the enzyme can use the product of a previous catalytic event as a substrate for a subsequent catalytic event. Interestingly, the membrane-permeabilized galacturonosyltransferase activity reported from pumpkin may represent a processive mode of elongation since reactions containing approximately equimolar amounts of UDP-GalA and acceptor yielded a population OGAs elongated by up

Table 5.4 Comparison of catalytic constants and pH optimum of HG- α 1,4-GalATs^{a,b}

Enzyme ^b	Plant source	Apparent K_m for		V_{max} (pmol mg ⁻¹ min ⁻¹)	Ref.
		UDP-GalA (μ M)	pH optimum		
GalAT ^a	Mung bean	1.7	6.0	~ 4700	(307)
GalAT	Mung bean	n.d. ^c	n.d.	n.d.	(322)
GalAT	Pea	n.d. ^c	6.0	n.d.	(313)
GalAT	Pea	n.d.	n.d.	n.d.	(230)
GalAT	Sycamore	770	n.d.	?	(309)
GalAT	Tobacco	8.9	7.8	150	(310)
GalAT (sol) ^d	Tobacco	37	6.3–7.8	290	(311)
GalAT (per) ^e	Azuki bean	140	6.8–7.8	2700	(314)
GalAT (sol)	Petunia	170	7.0	480	(315)
GalAT (per)	Pumpkin	700	6.8–7.3	7000	(319)
GalAT (sol)	Arabidopsis	n.d.	n.d.	n.d.	(316) (137)

^a Adapted from Refs. (192, 205).

^b Unless indicated, all enzymes are measured in particulate preparations.

^c n.d., not determined.

^d (sol): detergent-solubilized enzyme.

^e (per): detergent-permeabilized enzyme.

to five galacturonosyl residues (319). However, further information on the size distribution of the OGAs produced in reactions containing excess acceptor will be required to confirm this. Solubilized petunia galacturonosyltransferase, in reactions containing ~60-fold excess UDP-GalA to OGA acceptor, added up to 27 galacturonosyl residues onto the OGA acceptors (315), indicating that the enzyme can elongate OGA products from a previous reaction, but not specifically addressing the mode of elongation of the enzyme (320). The apparent lack of in vitro processivity of the solubilized GalAT suggests either that enzyme does not synthesize HG in a processive manner in vivo, or that the characteristics of HG-GalAT measured in vitro may be an artifact due to the dissociation of a required biosynthetic complex or loss of cofactor(s) or substrate(s) during solubilization of the enzyme. We have been unable to obtain evidence for processive in vitro solubilized GalAT activity (i.e., the production of extensively elongated acceptors under reaction conditions with excess OGA acceptor to UDP-GalA) (Quigley and Mohnen, unpublished results) [see (192)]. We also obtained no evidence that the inclusion of the methyl donor *S*-adenosylmethionine (308, 310, 314) and/or the acetyl donor acetylCoA promote the processivity of GalAT (192) (unpublished results). Clarification of the mode of action of GalAT and the mechanism of HG synthesis should be aided by access to purified or recombinantly expressed enzyme(s), and may require isolation of enzyme complexes (see below).

Efforts to purify GalAT activity from tobacco proved unsuccessful due to loss of activity. Therefore, a partial purification and tandem mass spectrometry approach was used to identify a gene encoding GalAT activity. The detergent-solubilized HG:GalAT from Arabidopsis suspension cells was partially purified by column chromatography to yield an enriched fraction containing approximately 20 protein bands. The proteins were trypsinized, the polypeptides analyzed for amino acid sequence by tandem mass spectrometry (137), and the sequences compared against the Arabidopsis gene database. The partially purified active

GalAT fraction contained two proteins with sequences that identified them as putative glycosyltransferases which were eventually named GAUT1 and GAUT7 (137). Evaluation of their amino acid sequences indicated that both GAUT1 and GAUT7 had characteristics consistent with the biochemical properties of HG:GalAT: a basic PI and an apparent Type II membrane protein topology. The encoded proteins consisted of three domains: a short N-terminal region, a single membrane spanning region, and a larger C-terminal domain. Transient expression of N-terminal truncated forms of the coding regions of GAUT1 and GAUT7 in human HEK293 cells indicated that GAUT1 had GalAT activity. Thus, GAUT1 became the first enzymatically verified HG:galacturonosyltransferase (*GalacatUronosylT*ransferase 1, GAUT1). The transiently expressed truncated form of GAUT7 did not have GalAT activity.

Sequence comparison of GAUT1 against the Arabidopsis genome identified 24 additional Arabidopsis genes with high sequence similarity to GAUT1. We named this 25-member group of related genes the GAUT1-related gene family (Table 5.5). The GAUT1-related genes represent a subclass of the Arabidopsis CAZy family GT-8 genes. Fifteen of the GAUT1-related genes have high sequence similarity to GAUT1 (37–100% identity/56–100% similarity) and we named these genes GAUT1–GAUT15. The 15 GAUT proteins have predicted masses of 61–78 kDa and most encode proteins with a predicted membrane anchor (GAUTs 1, 6–15) or with a signal peptide (GAUTs 3–5) consistent with a Type II membrane topology or with an intramicrosomal membrane location, respectively. GAUT2 is the only GAUT not predicted to be present in, or pass through, the intracellular membrane transport system (i.e., the ER/Golgi system). The remaining 10 GAUT1-related genes have somewhat lower sequence similarity to GAUT1 (39–44% identity/43–53% similarity) and we named these the GAUT1-Like (GATL) genes. The GATL genes encode predicted 33–44 kDa proteins with predicted signal peptides. The proven location of GAUT1 and GAUTs 3, 7, 8, and 9 is in the Golgi (323, 324).

Multiple T-DNA insert mutants of many of the GAUT and GATL genes are available (e.g., see <http://signal.salk.edu/cgi-bin/tdnaexpress>) and several mutants have described phenotypes. The *qual1/gaut8* mutant has ~25% reduced amounts of GalA in the walls of Arabidopsis rosette leaves or total plants (135), ~30% reduced GalA levels in walls of stem (260), and modest reductions in the GalA content of suspension cultured cells (325), suggestive of a role of the mutated gene as a putative GalAT involved in pectin synthesis. However, the mutant walls are also reduced in Xyl (at least stem walls) and protein extracts from mutant stems have both reduced HG:α-1,4-GalAT activity and β-1,4-xylosyltransferase activity (260). The lack of a confirmed enzyme activity of the recombinantly expressed or purified GAUT8 protein, along with the pleiotropic effects of the mutant, have made a conclusive identification of the function of the GAUT8 protein elusive.

The *parvus/glz1/gatl1* mutants (136, 326) also have characteristics consistent with a defect in pectin synthesis. The *parvus/glz1/gatl1* mutants grown under low humidity are semi-sterile dwarfs that have reduced anther dehiscence. The *parvus* mutants also have slightly elevated Rha, Ara, and Gal and reduced Xyl compared to WT (136). These changes in neutral sugar compositions are consistent with a role of the *parvus* gene in the synthesis of the pectic polysaccharide RG-I. However, since the levels of GalA in the mutant walls were not determined, it is not known if the mutant walls are altered in GalA content.

We PCR-tested 57 Arabidopsis GAUT1-related gene family T-DNA insertion mutant lines from the T-DNA mutant collection (<http://signal.salk.edu/cgi-bin/tdnaexpress>) and identified 36 homozygous mutant lines (Caffall and Mohnen, unpublished). Glycosyl residue

Table 5.5 The arabidopsis GAUT1-related gene superfamily

Gene ^a	Accession no. ^b	Amino acid identity/similarity to GAUT1 ^c	Clade	Enzyme activity confirmed	Putative enzyme activity	Mutants
<i>GAUT1/JS36/LGT1</i>	At3g61130	100/100	GAUT-A	HG:GalAT		
<i>GAUT2/LGT2</i>	At2g46480	65/78	GAUT-A			
<i>GAUT3</i>	At4g38270	68/84	GAUT-A			
<i>GAUT4/JS36L/LGT3</i>	At5g47780	66/83	GAUT-A			
<i>GAUT5/LGT5</i>	At2g30575	45/67	GAUT-A			
<i>GAUT6</i>	At1g06780	46/64	GAUT-A			
<i>GAUT7/JS33/LGT7</i>	At2g38650	36/59	GAUT-A			
<i>GAUT8/QUA1</i>	At3g25140	58/77	GAUT-B		Put. HG:GalAT and/or Put. β 1,4-XylIT	<i>qua1</i>
<i>GAUT9</i>	At3g02350	57/76	GAUT-B			
<i>GAUT10/LGT4</i>	At2g20810	50/72	GAUT-B			
<i>GAUT11</i>	At1g18580	51/71	GAUT-B			
<i>GAUT12/LGT6/IRX8</i>	At5g54690	40/61	GAUT-C		Put. HG:GalAT or Put. xylan primer GalAT	<i>irx8</i>
<i>GAUT13</i>	At3g01040	43/62	GAUT-C			
<i>GAUT14</i>	At5g15470	43/62	GAUT-C			
<i>GAUT15</i>	At3g58790	37/56	GAUT-C			
<i>GATL1/PARVUS/GLZ1</i>	At1g19300	29/49	GATL		<i>parvus/glz1</i>	
<i>GATL2</i>	At3g50760	27/52	GATL			
<i>GATL3</i>	At1g13250	23/43	GATL			
<i>GATL4</i>	At3g06260	29/51	GATL			
<i>GATL5</i>	At1g02720	25/44	GATL			
<i>GATL6/LGT10</i>	At4g02130	29/52	GATL			
<i>GATL7</i>	At3g62660	29/51	GATL			
<i>GATL8/LGT9</i>	At1g24170	23/42	GATL			
<i>GATL9/LGT8</i>	At1g70090	27/48	GATL			
<i>GATL10</i>	At3g28340	28/53	GATL			

^a The name given to each member of the GAUT1-related gene family includes its designation within the LGT family (332) and the names of any characterized Arabidopsis gene mutants (73, 74, 129, 132, 135, 136, 326). The numbering of the GAUT and GATL genes is based on the phylogenetic analysis of the family [see (137)].

^b From the Arabidopsis Information Resource database or NCBI.

^c Sequence identity/similarity is compared to 397 amino acids of GAUT1 starting at amino acid position 277.

composition analyses were carried out on amylase-treated walls from 82 unique tissue samples from homozygous mutants of 13 of the 15 GAUT family members by combined gas chromatography/mass spectrometry (GC/MS) of per-*O*-trimethylsilyl (TMS) derivatives of the monosaccharide methyl glycosides (327). Tissues from mutant lines of 8 of the 13 GAUT1-related family genes tested had $\geq 15\%$ reduction in the amount of galacturonic acid

in their walls, consistent with a function of the mutated genes as GalATs. However, a detailed biochemical analysis of the enzyme activity of each of the GAUTs will be required to establish gene function. Interestingly, some mutants also accumulated higher amounts of arabinose, rhamnose, fucose, and xylose and reduced amounts of mannose, galactose and/or glucose. The significance of these correlative changes with reduced levels of GalA is not yet clear; however, we speculate that the reduction of one pectic polysaccharide (e.g., HG) may be associated with, or compensated for, by a change in the amount of another polysaccharide (e.g., loss of HG may lead to an increase in RG-I which has Rha, Ara and/or Gal; alternatively, loss of an RG-I biosynthetic GalAT could lead to loss of RG-I and to an increase in HG with a resulting increase in GalA and a reduction in Rha, Ara and and/or Gal). Such a compensation of one wall polymer for another (328, 329) could occur via several mechanisms including sensing of polysaccharide or nucleotide-sugar levels, or the covalent or non-covalent association of one type of polysaccharide with another (e.g., it has been reported that pectin may be covalently linked to the hemicellulose xyloglucan (107)).

Although GAUT1 encodes a GalAT, its function in pectin synthesis is far from clear. For example, it is not yet understood at what stage of pectin synthesis, i.e. initiation or elongation, GAUT1 has its primary role. Also, it is not clear whether GAUT1 functions alone or in a complex. Our preliminary results suggest that, at least in vitro, GAUT1 can function in a complex with at least one other GAUT (Atmodjo and Mohnen, unpublished). The enzyme function of the other GAUT and GATL members of the GAUT1-related gene family remains to be determined. QUA1/GAUT8 is a good candidate for a GalAT, but the reduced levels of Xyl and xylosyltransferase activity of the *qual* mutants, at least in stem tissue, raises the question of what the connection is between pectin (e.g., HG) and hemicellulose (i.e., xylan) synthesis). The pectins and hemicelluloses are traditionally considered to be two different classes of wall polysaccharides, although there is some evidence that these two classes may be tightly, and possibly covalently, linked in the wall (258, 332, 333). Thus, the characteristics of the *qual* mutant raise several questions. Is xylan synthesis dependent upon pectin synthesis? The characteristics of the *parvus* mutant and the GATL proteins are also intriguing. What is the function of the GATLs? The characteristics of the *parvus* mutant are consistent with a function in pectin synthesis; however, these proteins have no apparent membrane spanning domain but rather a signal peptide. If the GATLs are involved in pectin synthesis, do they interact with other luminal pectin biosynthetic enzymes in the form of Golgi-localized complexes? The identification of GAUT1 as a functional GalAT and of the GAUT1-related gene family provides the gene/protein tools required to address some these questions.

Recently, characterization of mutants of GAUT12/IRX8 have provided evidence supporting a possible role of GAUT12/IRX8 as a HG:GalAT involved in the synthesis of a subfraction of HG to which β -1,4-xylan is attached (132) or as a putative α 1,4-GalAT that adds a GalA onto a xylose at the reducing end of a xylan primer (129). Proof of the function of GAUT12, however, requires confirmation of enzyme activity.

5.4.6.2 HG-methyltransferase

HG-methyltransferase (HG-MT) catalyzes the methylesterification of HG at the C-6 carboxyl group of some GalA residues by transfer of a methyl group from S-adenosylmethionine to HG. The name HG-MT is preferred, rather than the former name pectin

methyltransferase, to distinguish HG-MT from the enzymes that methylate RG-I or RG-II. HG-MT activity has been identified in microsomal membranes from mung bean (280, 281, 283), flax (227, 282), tobacco (284), and soybean (285). Membrane-bound HG-MTs from flax (333, 334), and tobacco (335) could be solubilized using detergent. HG-MT has been localized to the Golgi (227–229, 336) and its catalytic site has been shown to face the Golgi lumen (229). In vitro biochemical studies showing that UDP-GalA stimulates HG-MT activity in intact membrane vesicles (284, 308) and that polygalacturonic and pectin can function in vitro as acceptors for HG-MT in detergent-permeabilized membranes support a model in which at least a small stretch of HG is synthesized prior to its methylation by HG-MT in the Golgi. The observation that some of the HG-MTs in detergent-permeabilized membranes from flax and soybean show a preference for partially esterified pectin (228, 285, 337) over polygalacturonic acid further suggest that multiple HG-MTs may exist that differ in their specificity for HG of differing degrees of methylation. The question of whether such HG-MTs are preferentially involved in the initial methylation of HG or in the methylation of more highly esterified HG remains to be resolved. The observation that overexpression of an Arabidopsis S-adenosylmethionine (SAM) synthetase gene in flax leads to a concomitant increase in SAM synthetase activity and of pectin methylesterification in the wall, with no increase in HG-MT activity, suggests that the degree of methylesterification of HG may be regulated, at least in part, at the level of substrate (i.e., SAM) concentration (338).

The gene encoding HG-MT has not yet been unambiguously identified. Two apparent HG-MT isozymes, PMT5 and PMT7, were reported from flax (337) and efforts to purify these apparent isozymes resulted in the identification of an additional small polypeptide with HG-MT activity designated PMT18. The definitive identification of one or more of these polypeptides as HG-MT has not yet been reported. Thus, the proposition that the 18-kDa protein is a subunit of the 40- and 110-kDa proteins (337) has not been substantiated.

Recently, Mouille and colleagues (287) have identified a putative methyltransferase as the gene mutated in the Arabidopsis mutant *quasimodo2*. *Qua2-1* plants are dwarfed and have a 50% reduction in HG. Although confirmation of enzyme activity of QUA2 is required to establish if it indeed encodes an HG-MT, the reduced HG phenotype of *qua2* plants along with the Golgi localization of QUA2–GFP fusions and the putative methyltransferase domain in QUA2 are consistent with a role as an HG-MT. Further work on QUA2 and the 29 QUA2-related proteins in Arabidopsis may shed light on the identity of multiple methyltransferases required for pectin synthesis.

5.4.6.3 HG-acetyltransferase (HG-AT)

HG may be partially O-acetylated at C-2 or C-3 of GalA (198, 199). No gene for HG acetyltransferase has been identified; however, O-acetyltransferase activity in microsomes from suspension-cultured potato cells (339) has been shown to transfer [14 C]acetate from [14 C]acetyl-CoA onto endogenous acceptor in the microsomes to yield a salt/ethanol precipitable product from which approximately 8% of the radioactivity could be solubilized by treatment with endopolygalacturonase and pectin methylesterase. Such results could indicate the presence of HG-AT, although the possibility that the radiolabeled acetate was transferred either onto RG-II or RG-I that was solubilized by the glycanase treatments, and thus represents an enzyme that acetylates one of the other pectic polysaccharides that may be covalently linked to HGA, cannot be ruled out.

5.4.7 Xylogalacturonan synthesis

HG may contain regions that are substituted with β -D-xylose linked to C-3 of GalA (185, 186, 193, 267). Such regions of xylosylated HG are referred to as xylogalacturonan (XGA) and have been most frequently identified in reproductive tissues of plants including apple (184, 193, 197), cotton and watermelon (185), and pea (342), but also in carrot (186). However, xylogalacturonan has also been detected in *Arabidopsis* leaves and stems (187), albeit it in lower levels than in reproductive and storage tissues such as apple and potato. No gene for XGA:xylosyltransferase (XGA:XylT) has been unambiguously identified. XGA:XylT activity was identified in studies of apiogalacturonan synthesis (341, 342). Although the product produced was not characterized in detail, at least some of the radioactive xylose appeared to be incorporated into apiogalacturonan and/or HG.

Interestingly, Nakamura and coworkers (263) has reported that in soybean some XGA may be further elongated with β -1,4-linked xylose residues yielding β -1,4-xylans of up to seven xylosyl residues in length. Such results suggest that HG may, at least in soybean, be a primer or acceptor for a glycosyltransferase or a transglycosylase that establishes a link between pectin and the hemicellulose xylan. Such a linkage would be consistent with the characteristics of the *Qual* mutant (mutated in GAUT8) (260) and the *irx8* mutant (mutated in GAUT 12) (132).

5.4.8 Apiogalacturonan synthesis

5.4.8.1 Apiogalacturonan-galacturonosyltransferase (AP:GalAT)

It is not known whether apiogalacturonan is synthesized on preexisting HG that is synthesized by GAUT1 or related GalATs, or whether a unique GalAT is responsible for apiogalacturonan synthesis. There have been no reports of efforts specifically targeted at identifying the apiogalacturonan:GalAT.

5.4.8.2 Apiogalacturonan-apiosyltransferase (AP:ApiT)

Apiogalacturonan is a substituted galacturonan that is produced in some aquatic monocotyledonous plants (188, 189) and that consists of HG substituted at O-2 or O-3 with apiose or apiobiose (D-Apif- β -1,3-D-apiose) (188, 189). The anomeric configuration of the linkage of apiose to HG may be in the β configuration (189). It is not known whether the same apiogalacturonan:ApiTs synthesize RG-II (see below) and apiogalacturonan. For example, RG-II has two of its four side branches attached to an HG backbone by a β -Apif linked to the O-2 of HG (158), and thus, the possibility exists that the β 1,2-apiosyltransferase involved in RG-II synthesis may also synthesize apiogalacturonan. In vivo synthesis of apiogalacturonan has been studied in vegetative fronds of *Spirodela polyrrhiza* (343) and D-apiosyltransferase activity has been characterized in cell-free particulate preparations from duckweed (*Lemna minor*) (341). The apiosyltransferase in particulate membrane preparations from *Lemna* transfers [14 C]-apiose from UDP-[14 C]-apiose onto endogenous acceptors. The enzyme has an apparent K_m for UDP-apiose of 4.9 μ M and a pH optimum of 5.7 (341). Since, the rate of apiosyltransferase activity increased twofold when UDP-GalA was added to the reaction

(341) and the product synthesized in the presence of UDP-GalA bound anion exchange resin more tightly than the product synthesized without UDP-GalA (342), it is likely that the apiosyltransferase transfers apiose onto a growing HG chain. The ApiT has not been purified and the gene has not been identified.

5.4.9 Synthesis of rhamnogalacturonan II (RG-II)

RG-II is the most structurally complicated polysaccharide in the cell wall. It is present in the walls of all plants and its structure is highly conserved. RG-II makes up approximately 4% of the cell wall in dicotyledonous plants and less than 1% of the wall in monocots (157). It contains 12 different types of glycosyl residues in at least 20 different linkages (159) including both methyl etherified (e.g., 2-*O* Me-xylose and 2-*O* Me-fucose) and *O*-acetylated glycosyl residues (e.g., 3-*O*-, or 4-*O*-Ac-fucose). RG-II also contains unusual sugars such as aceric acid (3-*C*-carboxy-5-deoxy-*L*-xylose) (346), KDO (2-keto-3-deoxy-*D*-manno-octulopyranosylonic acid) (345), and DHA (3-deoxy-*D*-lyxo-2-heptulopyranosylaric acid) (346). RG-II has a backbone of α -1,4-linked *D*-galactosyluronic acid with structurally complex side chains attached to C-2 and/or C-3 (2, 157, 159, 269, 345–352). RG-II in the wall exists largely complexed with borate as an RG-II dimer that is cross-linked by a borate diester (159, 352–357). At least 24 transferase activities are expected to be required for RG-II synthesis. There have been very few systematic studies of RG-II synthesis, although progress is beginning to be made on several enzymes.

5.4.9.1 RG-II:galacturonosyltransferase (RG-II:GalAT)

It is not known whether RG-II is synthesized on preexisting HG that is synthesized by GAUT1 or related GalATs, or whether a unique GalAT is responsible for synthesizing the HG backbone of RG-II. It is also not known if the side chains of RG-II are synthesized as individual oligosaccharides and transferred in bulk onto HG, or whether individual glycosyltransferases transfer each distinct glycosyl residue individually onto the growing non-reducing end of the maturing RG-II molecule. The results of Egelund *et al.* (274) on the proposed RG-II:XylTs would be in agreement with the latter model (see below).

5.4.9.2 RG-II:apiosyltransferase (RG-II:ApiT)

As mentioned above it is possible that the apiosyltransferase activity identified during the studies of apiogalacturonan synthesis in *Lemna* is involved in RG-II synthesis. However, no gene for RG-II ApiT has been identified. Interestingly, when the gene reported to encode the UDP-apiose or UDP-*D*-apiose/UDP-*D*-xylose synthase was downregulated in *Nicotiana benthamiana* by virus-induced gene silencing of *NbAXS1*, the result was a reduction in the amount of RG-II in the walls (358). These results provide evidence that UDP-apiose is the substrate for incorporation of apiose into RG-II and that the UDP-*D*-apiose/UDP-*D*-xylose synthase gene encodes the enzyme that synthesizes the required substrate for RG-II synthesis.

5.4.9.3 *RG-II:xylosyltransferase (RG-II:XylT)*

Recently, work from the Ulvskov and Geshi groups (274) has provided strong evidence that these investigators have identified two *Arabidopsis thaliana* RG-II- α -D-1,3-xylosyltransferases (RG-II- α 1,3XylTs) (274). Following the identification of a novel family of 27 putative *Arabidopsis thaliana* glycosyltransferases (215) and through a series of bioinformatic analyses aimed at identifying novel plant cell wall biosynthetic glycosyltransferases with a predicted Type II membrane topology (359), two of these genes, named *RGXT1* (At4g01770) and *RGXT2* (At4g01750) were shown to encode proteins with characteristics consistent with a function as RG-II- α 1,3XylTs. *RGXT1* and *RGXT2* encode proteins of 361 and 367 amino acids, respectively, share 90% sequence identity, and are members of GT-family 77 (138) (<http://afmb.cnrs-mrs.fr/CAZY/>). Two additional *Arabidopsis* genes, At4g01220 and At1g56550, are 68–75% identical to *RGXT1* and *RGXT2* (274). Expression of truncated soluble forms of *RGXT1* and *RGXT2* in baculovirus-transfected insect cells and enzyme assays using diverse radiolabeled nucleotide-sugars and free monosaccharide acceptors demonstrated that the expressed proteins catalyze the transfer of Xyl from UDP- α -D-Xyl onto fucose. Biochemical analyses of the synthesized product using specific xylosidases and NMR spectroscopy indicated that the xylose was transferred onto the fucose in an α -1,3-linkage. Based on these results the authors hypothesized that *RGXT1* and *RGXT2* function in the synthesis of the RG-II side chain A that contains 2-*O*-methyl-D-Xyl attached in an α 1,3-linkage to α -L-Fuc. Acceptor specificity studies demonstrated that both enzymes preferred L-Fuc with an α -anomeric linkage and disaccharide acceptors with Fuc attached at the position 4, rather than at the 2 or 3 position, to another glycosyl residue; all characteristics consistent with the structure of RG-II (159). Importantly, RG-II isolated from *RGXT1* and *RGXT2* mutant walls, but not RG-II from wild type *Arabidopsis* walls, served as an acceptor for the enzyme, providing strong evidence that *RGXT1* and *RGXT2* function in RG-II synthesis (274) and providing strong support for the function of *RGXT1* and *RGXT2* as RG-II- α 1,3XylTs. The lack of a clear difference in the structure of RG-II isolated from walls of *RGXT1* and *RGXT2* mutants compared to wild type walls, however, is perplexing and leaves open the question of whether there is gene redundancy, thus requiring a double (or more) gene knockout mutant to see a phenotype. Alternatively, the question remains as to whether *RGXT1* and *RGXT2* may have additional or alternative functions in the synthesis of some other, yet to be identified, wall polysaccharide structure. Further studies of *RGXT1* and *RGXT2* and related genes should clarify their role(s) in pectin synthesis.

5.4.9.4 *RG-II:glucuronosyltransferase (RG-II:GlcAT)*

Studies of the *Nicotiana plumbaginifolia* T-DNA nolac-H18 callus mutant lead to the identification of the mutated gene, *NpGUT1*. *NpGUT1* has 60% sequence homology to animal glucuronosyltransferases that synthesize heparin sulfate. Complementation of the nolac-H18 mutant with the *NpGUT1* gene corrected the non-organogenesis and weak intercellular attachment phenotypes of the mutant. Cell walls of the nolac-H18 mutant contained 86% reduced levels of glucuronic acid, a reduction that was associated with the pectin-enriched fraction of the walls. RG-II from mutant walls was devoid of glucuronic acid, leading the authors to propose that *NpGUT1* encodes RG-II- β -1,2GlcAT that transfers GlcA onto the L-fucose in RG-II side chain A (181). The mutant RG-II in the nolac-H18 showed 82% reduced

RG-II dimer formation, providing further support that RG-II is modified in the mutant. Taken together these results show that *NpGUT1* encodes a putative RG-II:β-1,2GlcAT. Enzymatic confirmation of the activity of the encoded protein has not yet been reported.

5.4.9.5 *RG-II:methyltransferase (RG-II:MT)*

RG-II contains 2-*O*-methylfucose and 2-*O*-methylxylose. It is not known if the methyl group is added at the stage of the nucleotide-sugar or after the sugar is added to RG-II. The genes encoding the methyltransferases have not been identified.

A pectin methyltransferase activity detergent-solubilized from suspension-cultured flax cells was able to transfer methyl groups from *S*-adenosylmethionine onto RG-II isolated from wine (334). Enzyme reactions containing RG-II had sevenfold great methyltransferase activity than reactions without exogenous acceptor and the radiolabeled product synthesized had a size similar to RG-II monomers and RG-II dimers (158, 334). It was not established where in RG-II the methyl group was added and thus, the methylation may have represented methylesterification of the HG backbone of RG-II, or alternatively, could have been due to methyletherification of RG-II since RG-II contains methyl groups on non-galacturonic glycosyl residues (e.g., 2-*O*-methyl xylose and 2-*O*-methyl fucose (352, 360) of side chain residues. The location of the methylation in RG-II and the identity of the potentially novel enzyme activity that catalyzes its incorporation into RG-II have not been established.

5.4.9.6 *RG-II:acetyltransferase (RG-II:AT)*

RG-II contains 3-*O*-acetylacetic acid and acetylated methyl fucose. No gene for RG-II acetyltransferase has been identified. However, *O*-acetyltransferase activity in microsomes from suspension-cultured potato cells (339) has been shown to transfer [¹⁴C]acetate from [¹⁴C]acetyl-CoA onto endogenous acceptors in the microsomes to yield a salt/ethanol precipitable product from which approximately 8% of the radioactivity could be solubilized by treatment with endopolygalacturonase and pectin methylesterase. Thus, it is possible that the radiolabeled acetate solubilized by the glycanase treatments represented acetylated RG-II and that the activity identified was RG-II acetyltransferase.

5.4.9.7 *Other RG-II transferases*

The glycosyltransferases that insert fucose, KDO, DHA, and acetic acid into RG-II have not been identified. A 10-member Arabidopsis gene family (118) that has 35–73.8% amino acid sequence identity to an Arabidopsis α1,2-fucosyltransferase that fucosylates a side branch in the hemicellulose xyloglucan has been described (116). Whether one or more of these genes encodes fucosyltransferase(s) involved in RG-II synthesis remains to be investigated.

5.4.10 *Rhamnogalacturonan I (RG-I) synthesis*

RG-I accounts for 7–14% of the primary wall (157) and 20–33% of pectin (194). Unlike HG and RG-II, RG-I has a backbone of up to 100 repeats of the disaccharide [→4)-α-D-GalpA-(1→2)-α-L-Rhap-(1→] (2, 157, 261, 262, 361, 362). The GalA residues in RG-I

may be *O*-acetylated at C-3 or C-2 (157, 199, 288–290). The average molecular weight of sycamore RG-I has been reported to be 10^5 – 10^6 Da (157). Between 20 and 80% of the rhamnosyl residues are substituted at C-4, and sometimes at C-3, with side chains composed mostly of arabinosyl and/or galactosyl residues (2, 157, 264), referred to as galactans (157, 264, 278), arabinans (214, 264, 290), and arabinogalactans (2, 157, 190, 214, 264). These side chains may range in size from 1 to 50 or more glycosyl residues (2, 157, 290). A large amount of immunocytochemical evidence based on antibodies against specific RG-I carbohydrate epitopes (160) indicates that the precise structure of the side chains of RG-I varies in a cell type and development-specific manner (214, 363). Representative side chains or portions of side chains that have been identified in RG-I have been previously summarized in (192).

The RG-I galactans may contain only galactosyl residues or may contain other neutral glycosyl residues (157) or acidic residues such as GalA (190), GlcA (157, 190, 278), or 4-*O*-methylGlcA (278). Some galactans also have β -1,6-branching (190). As mentioned above, the size and linkages in the galactan side branches of RG-I vary depending upon the species (157).

RG-I also contains side chains of individual or multiple *L*-arabinofuranosyl (Ara *f*) residues or chains of 1,5-linked- α -*L*-Ara *f* substituted at O-3 and occasionally at O-2 with additional Ara *f* residues (190, 290, 364). Such side chains are referred to as arabinans.

Some RG-I side chains contain both arabinosyl and galactosyl residues. These side chains are referred to as arabinogalactans that have been divided into Type I and Type II arabinogalactans. Type I arabinogalactans contain a 1→4-linked β -D-galactan backbone while the Type II arabinogalactans contain a 1→3-linked β -D-galactan backbone and are often highly branched (2, 157, 190, 214). Type II arabinogalactans may be associated with glucuronomannoglycans (190). Some studies suggest that mannose may be a component in some pectins, probably as a side branch to RG-I (190), however, the structural role of mannose in pectin has not been clearly demonstrated and therefore mannose is not discussed as a component in pectin here. Some of the Type II arabinogalactan is associated with arabinogalactan proteins (AGPs) (365–368), hydroxyproline-rich proteins located at the plasma membrane, cell wall, or in media surrounding suspension-cultured cells (366, 367, 369, 370). It is not always clear whether specific Type II arabinogalactan structures isolated from wall extracts are associated with RG-I, AGPs, or both. However, multiple lines of evidence show that some Type II arabinogalactan is associated with RG-I. This includes the cross reactivity of the antibody CCRC-M7 with both RG-I and arabinogalactan proteins (371). CCRC-M7 recognizes a trimer or larger of β -(1,6)-Gal carrying one or more Ara residues (372). Pectic polysaccharides from the *Chenopodiaceae* family including spinach (*Spinacia oleracea*) and sugar beet (*Beta vulgaris*) are esterified with phenolics such as ferulic acid (157, 291, 373), on galactose and arabinose residues that are likely substituents in the side branches of RG-I (157, 291, 292, 374).

5.4.10.1 RG-I:GalAT, RG-I:RhaT, RG-I:GalAT/RhaAT

It is not known whether RG-I is synthesized onto existing HG or rather is synthesized independent of HG. If it is synthesized onto HG, it not known whether GAUT1 is that GalAT responsible for synthesizing the HG backbone region which would serve as a primer for RG-I backbone synthesis, or whether other GAUTs or other enzymes would perform this function.

It is also not known whether the alternating $[\rightarrow 4)\text{-}\alpha\text{-D-GalpA-(1}\rightarrow 2)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow]$ backbone repeat is synthesized by a single glycosyltransferase containing both RhaT and GalAT activity, or whether, alternatively, the RG-I backbone is synthesized by a protein complex containing both a GalAT and a RhaT. If the backbone is synthesized by a complex, it is also not known whether GAUT1 or one or more of the GAUT1-related gene family members are part of the complex. To date, no RG-I-specific GalAT or RhaT has been reported.

5.4.10.2 RG-I:galactosyltransferase (RG-I:GalT)

RG-I synthesis requires at least eight galactosyltransferase (GalT) activities (Table 5.2). Probable β -1,4-GalT and β -1,3-GalT activities were originally identified in studies of microsomal preparations from mung bean (375, 376) and more recently a mung bean β -1,4-galactosyltransferase activity with a pH optimum of 6.5 was confirmed (377). Multiple galactosyltransferase (GalTs) activities have also been reported in particulate homogenates (378, 379) and solubilized enzyme (380) from flax (*Linum usitatissimum* L.). Detergent-solubilized flax microsomal GalTs transferred $[^3\text{H}]\text{Gal}$ from UDP- $[^3\text{H}]\text{Gal}$ onto exogenous RG-I-enriched and pectic β -1,4-galactan acceptors (381) to yield high molecular mass radiolabeled products. Surprisingly, the pH optimum for transfer onto lupin pectic β 1,4-galactan (i.e., pH 6.5) was different than the pH optimum for transfer of Gal onto an endopolygalacturonase-treated RG-I-enriched fractions from flax (i.e., two optima: pH 6.5 and 8.0) (381). Analysis of the products using RG-I-specific enzymes confirmed that the GalTs indeed added Gal onto RG-I (381), and thus, represented RG-I:GalTs. Furthermore, fragmentation of at least part of the product with β -1,4-endogalactanase demonstrated that at least some of the GalT activity represented β -1,4-galactosyltransferase (381). At pH 8, the GalT activity had an apparent K_m of 460 μM for UDP-Gal and characteristics consistent with a function in catalyzing the addition of galactose onto short galactan side branches of RG-I.

Microsomal membranes from potato suspension cultured cells have been shown to contain RG-I: β -1,4-galactosyltransferase activities that both initiate and elongate β -1,4-galactan side chains of RG-I (382). The potato RG-I: β -1,4-GalT activity in microsomal membranes had a pH optimum of 6.0–6.5 and produced a >500 kDa product using UDP- $[^{14}\text{C}]\text{Gal}$ and endogenous acceptor(s) in microsomal membranes. The product was fragmented by endo- β -1,4-galactanase into $[^{14}\text{C}]\text{Gal}$ and $[^{14}\text{C}]\text{galactobiose}$ and into radiolabeled fragments between 50 and 180 kDa in size (382) when treated with the RG-I-specific rhamnogalacturonase A, an endohydrolase that cleaves the glycosidic linkage between the GalA and Rha in the RG-I backbone (383). The GalT activities in the microsomal membranes could be solubilized from the membranes using detergent (382) and the solubilized enzyme fraction was shown to contain at least two distinct GalT activities, one with a pH optimum of 5.6 that preferentially added Gal onto an $\sim 1.2\text{-MDa}$ RG-I acceptor with a mole % Gal/Rha ratio of 0.7; and the other with a pH optimum of 7.5 that preferentially added Gal onto a 21-kDa RG-I acceptor with a mole % Gal/Rha ratio of 1.2. Neither activity could use RG-I acceptor containing lower Gal/Rha ratios, RG-I backbone without side chains, or galactan polymers or oligomers as acceptors, suggesting that the activities identified required recognition of the RG-I backbone and some existing Gal in a side chain (384). Interestingly, only the product synthesized onto the 21-kDa RG-I acceptor was digestible with a

1,4- β -endogalactase, suggesting that either the Gal transferred onto the larger RG-I acceptor was of a linkage other than β ,1-4, or that the length of the galactan side chain synthesized was less than three, the minimum size recognized by the 1,4- β -endogalactanase. The RG-I: β -1,4-GalT that elongates the β -1,4-side chains of RG-I was shown, by subcellular organelle fractionation and protease sensitivity experiments, to be a Golgi-resident protein with its catalytic site facing the lumen of the Golgi (385), a location consistent with its role in pectin synthesis. No gene for any RG-I galactosyltransferase has been reported.

More recently, β -1,4-GalT activity in mung bean detergent-treated microsomal membranes was identified that transferred up to eight galactosyl residues in a β -1,4-linkage onto the non-reducing end fluorescently labeled exogenous (1 \rightarrow 4)- β -galactooligosaccharide acceptors (386, 387). Of the galactooligosaccharide acceptors used, i.e., degree of polymerization (DP) of 3–7, the galactoheptaose was a most effective acceptor although acceptors of DP 4–6 also functioned. The fluorescently labeled trimer was not active. Interestingly, fluorescently labeled RG-I backbone oligosaccharides of DP 5–7 were also not active, suggesting that the GalT activity identified could not add Gal onto oligosaccharide RG-I backbone regions. The β -1,4-galactan: β -1,4-GalT activity had a pH optimum of 6.5 and apparent K_m of 32 μ M for UDP-Gal and 20 μ M for the fluorescently labeled galactoheptaose (386).

5.4.10.3 RG-I:arabinosyltransferase (RG-I:AraT)

RG-I contains L-arabinose in multiple linkages (see Table 5.2). Most of the arabinose is in the furanose ring form, although a terminal arabinose exists in the pyranose form in some RG-I side chains (268). AraT activity was originally identified in microsomes from mung bean (*Phaseolus aureus*) (388) and bean (*Phaseolus vulgaris*) (389) although definitive evidence that those AraT activities were involved in pectin synthesis was not demonstrated [see Ref. (205) for review]. The bean AraT activity was primarily associated with enriched Golgi, and to a lesser extent enriched endoplasmic reticulum (390).

The study of AraTs specifically involved in pectin synthesis has been problematic for several reasons. Multiple wall polysaccharides and proteoglycans contain arabinose, including pectin, hemicelluloses (e.g., glucuronoarabinoxylan), and arabinogalactan proteins. Thus, experiments aimed at studying pectin biosynthetic AraTs by incubating microsomal membranes with radiolabeled UDP-Ara have not been very successful. The nucleotide-sugar donor, UDP-Ara, was not available and had to be synthesized (391), although more recently the UDP- β -L-arabinopyranose form has become available through CarbSource (http://www.crc.uga.edu/~carbosource/CSS_home.html). However, while most Ara in pectin is in the furanose form, the nucleotide-sugar synthesized by the 4-epimerization of UDP- α -D-Xyl is UDP- β -L-arabinopyranose. Thus, this has been the nucleotide-sugar form most readily available for experimental use. However, there is uncertainty as to the nature of the nucleotide-sugar substrate used for pectin synthesis. Is it UDP- β -L-arabinopyranose (UDP-Arap) or UDP- β -L-arabinofuranose (UDP-Araf)? If it is UDP- β -L-arabinofuranose, how is this synthesized by the plant and, experimentally, what is the most facile way to produce it? Is it synthesized by enzyme-catalyzed ring contraction of UDP-L-arabinopyranose by a mutase (392)? Indeed, recently, Ishii and colleagues (393) identified a UDP-arabinopyranose mutase (UAM) from rice that catalyzes the reversible formation of UDP-Araf from UDP-Arap. Interestingly, UAMs are the same proteins that previously were identified as reversibly glycosylated polypeptides (RGPs), proteins that are reversibly

glycosylated in the presence of UDP-Glc and several other nucleotide-sugars (394–396). The significance of the reverse glycosylation detected *in vitro*, in regards to the role(s) of UAM in wall synthesis, remains to be determined.

Recent efforts to investigate the AraT activity in mung bean confronted some of the above-mentioned problems. Incubation of mung bean microsomal membranes with UDP- β -L-[14 C]arabinopyranose (UDP-[14 C]Ara) resulted in the incorporation of both [14 C]Ara and [14 C]Xyl into elongated endogenous acceptors because of the epimerization of some of the UDP-[14 C]Ara into UDP-[14 C]Xyl by a UDP-Xyl-4-epimerase present in the microsomal fraction (397). Furthermore, digestion of the radiolabeled product synthesized in the microsomes with endo-arabinase yielded very little radiolabeled Ara or arabinose-containing small oligosaccharides, suggesting that the conditions used were not conducive for the synthesis of arabinans. Conversely, digestion of the product with arabinofuranosidase did release some [14 C]Ara, indicating that an enzyme activity that could add at least a single Ara residue was present in the microsomes. A breakthrough in identifying pectin biosynthetic AraTs came when detergent-solubilized microsomal membranes and defined arabinooligosaccharides were used as acceptors for study of pectin biosynthetic AraTs (397). The incubation of detergent-solubilized microsomal membranes with (1 \rightarrow 5)- α -L-arabinooligosaccharides of DP 5–8 and with UDP- β -L-[14 C]arabinopyranose led to the identification of an AraT activity that could transfer a single arabinopyranose residue onto the non-reducing end of α 1,5-arabinooligosaccharide acceptors. The enzyme had a pH optimum of 6.5 and was shown to reside predominantly in the Golgi by subcellular fractionation of organelles (397). The anomeric configuration and linkage of the arabinopyranose residue transferred by the mung bean AraT was subsequently shown to be β -(1 \rightarrow 3) through the use of fluorescently labeled α -L-arabinooligosaccharide acceptors (271). Thus, mung bean contains an α 1,5-arabinan: β -(1 \rightarrow 3)arabinopyranose AraT (271).

A second mung bean arabinopyranosyltransferase activity was identified (273) that could transfer an individual arabinopyranosyl residue from UDP- β -L-[14 C]arabinopyranose onto the non-reducing end of fluorescently labeled 1,4-linked β -D-galactooligosaccharides of DP 3–7, with significantly better activity manifested with 1,4-linked β -D-galactooligosaccharides of DP 5 or greater. The β -1,4-galactan:AraT activity transferred the Ara residue in an α configuration on the O-4 position of the galactooligosaccharides, identifying the AraT as a β -1,4-galactan: α 1,4AraT. The enzyme had a pH optimum of 6.0–6.5 and apparent K_m (s) for UDP- β -L-[14 C]arabinopyranose and fluorescently labeled galactohexasaccharide of 330 μ M and 45 μ M, respectively. Interestingly, the enzyme would not use fluorescently labeled galactooligosaccharides of DP 6–10 that had a single α -L-Arap residue at the non-reducing end as acceptors for the previously described β -1,4-galactan: β -1,4-GalT (386), indicating that the enzyme cannot use mono-arabinosylated galactooligosaccharides as acceptors. The authors propose that the presence of the α -L-arabinopyranosyl residue on the β -1,4-galactan oligosaccharides prevents further galactosylation of the galactooligosaccharides (273).

Recently, a gene encoding a putative arabinan: α -1,5-arabinosyltransferase (*ARAD1*; At2g35100) has been identified in *Arabidopsis* (270) through analysis of the CAZy GT47 family glycosyltransferase gene At2g35100 (138) (<http://afmb.cnrs-mrs.fr/CAZY/>) and phenotypic biochemical and immunochemical analyses of the corresponding *Arabidopsis* T-DNA insert mutant *ARABINAN DEFICIENT 1*. *ARAD1* encodes a protein with a predicted molecular mass of 52.8 kDa and a single transmembrane helix region near the N-terminus.

Although homozygous knockout mutants of *ARAD1* show no visible growth differences from wild type, isolated walls from mutant leaves and stems had 25 and 54%, respectively, reduced levels of Ara compared to wild type walls (270). Transformation of the mutant plant with the *ARAD1* gene complemented the mutant phenotype (i.e. restored the amount of Ara in the wall to wild type levels), thus providing evidence that *ARAD1* affects wall arabinose levels. Immunocytochemical analysis of leaf, inflorescence stem, and stem revealed a reduction in immunolabeling with the anti- α -1,5-arabinan antibody LM6. A lack of difference between the labeling of protein extracts from wild type and mutant with LM6 suggested that the mutant was affected in the synthesis of α -1,5-arabinans, but not in glycoprotein synthesis (270). This observation was confirmed by comparison of RG-I isolated from wild type and *ARAD1* walls. Mutant RG-I had a 68% reduction in Ara content, which linkage data showed was predominantly due to a reduction in 5-linked Ara_f and also in 2,5 *f*-linked Ara and 2,3,5-linked Ara_f. These results strongly suggest that *ARAD1* is a putative RG-I arabinan: α -1,5-arabinosyltransferase. Confirmation of this activity will require expression of enzymatically active enzyme expressing α -1,5-arabinosyltransferase activity.

Recently, a novel approach to identify genes involved in cell wall synthesis has been taken (398) and offers promise in leading to the positional cloning for a gene that affects the number of arabinan side chains in RG-I. The method takes advantage of the natural variation that occurs in cell wall synthesis in natural plant populations and of the availability of Arabidopsis recombinant inbred line (RIL) populations which facilitate the identification and cloning of quantitative trait loci (QTLs). Through the use of multiple techniques to analyze cell walls of an RIL population from a cross between Arabidopsis Bay-0 and Shahdara, including global monosaccharide composition and Fourier-transform infrared (FTIR) microspectroscopy, a major QTL was identified that accounted for 51% of the heritable variation observed for the arabinose–rhamnose ratio in the cell walls, a difference that appeared to be due to variation in the amount of RG-linked arabinan. Whether this strategy will lead to the identification of RG-I biosynthetic AraTs remains to be shown.

Two Arabidopsis putative arabinosyltransferases, designated reduced residual arabinose-1 and -2 (RRA1; At1g75120 and RRA2; At1g75110) were recently identified by Egelund *et al.* (399) based on a 20% reduction in the arabinose content in pectin- and xyloglucan-depleted cell wall fractions from meristematic tissue of *rra1* and *rra2* mutant plants. However, whether these genes, which are classified in CAZy family GT77, encode functional arabinosyltransferases, and if so, whether they function in pectin, arabinoxylan, wall structural protein, or other syntheses, remains to be determined.

5.4.10.4 RG-I:methyltransferase (RG-I:MT)

Detergent-solubilized microsomal proteins from flax can use an RG-I-enriched fraction as an exogenous acceptor for methylation in the presence of S-adenosylmethione (334). The pectin methyltransferase activity was stimulated by the addition of the enriched RG-I fraction 1.5- to 1.7-fold above levels recovered using endogenous acceptor, and the radio-labeled product had a size similar to RG-I. However, since it was not shown where in RG-I the methylation occurred, it is not clear whether the methylation occurred on GalA in the RG-I backbone or, rather, on possible HG tails that may have been covalently linked to RG-I. Also, it was not established whether some of the methylation may have occurred on

a non-galacturonic substituent such as methylation at the 4-position of glucuronic acid in the side branched of RG-I (278).

5.4.10.5 *RG-I:acetyltransferase (RG-I:AT)*

GalA in the alternating $[\rightarrow 4)\text{-}\alpha\text{-D-GalpA-(1}\rightarrow 2)\text{-}\alpha\text{-L-Rhap-(1}\rightarrow]$ backbone of RG-I may be acetylated on C-2 and/or C-3 (289). Microsomal membranes from suspension-cultured potato cells (339) contain an RG-I acetyltransferase that transfers $[^{14}\text{C}]$ acetate from $[^{14}\text{C}]$ acetyl-CoA onto endogenous RG-I acceptor to yield a >500-kDa radiolabeled product (339), based on the release of $[^{14}\text{C}]$ acetate following incubation of the product with a purified rhamnogalacturonan O-acetyl esterase, and fragmentation of the product by rhamnogalacturonan lyase (RGase B) (339). The RG-I acetyltransferase has an apparent K_m for acetyl-CoA of 35 μM and a pH optimum of 7.0.

5.5 The cell biology and compartmentalization of cell wall synthesis

Except for cellulose and callose, all of the other plant polysaccharides appear to be synthesized in the Golgi or to pass through the Golgi en route to the cell wall. The available experimental results do not currently lead to an unambiguous picture of how the diverse wall polysaccharides are synthesized in, and travel through the Golgi to the wall. Immunoelectron microscopy using antibodies to XG, HG, and RGI showed that these polysaccharides are found in Golgi vesicles but not in the ER (106, 234). Different types of Golgi cisternae contain different sets of glycosyltransferases. Thus, the functional organization of the biosynthetic pathways of complex polysaccharides is consistent with these molecules being processed in a cis-to-trans direction like the N-linked glycans. RG-I and HG polysaccharides appear to be synthesized in cis- and medial-cisternae (235). Methylesterification of the carboxyl groups of the galacturonic acid residues in the polygalacturonic acid domains occurs mostly in medial cisternae, and arabinose-containing side chains of the polygalacturonic acid domains are added to the nascent polygalacturonic acid/rhamnogalacturonan-I molecules in the trans-cisternae. In root tip cortical parenchyma cells, anti-RG-I and the anti-XG antibodies are shown to bind to complementary subsets of Golgi cisternae, and several lines of indirect evidence suggest that these complex polysaccharides may also exit from different cisternae (224). On the other hand, xyloglucan and polygalacturonic acid/rhamnogalacturonan-I can be synthesized concomitantly within the same Golgi stack.

O-linked arabinosylation of the hydroxyproline residues of extensin occurs in cis-cisternae, and the glycosylated proteins pass through all cisternae before they are packed into secretory vesicles in the monensin-sensitive, trans-Golgi network (224). The β -1,4-linked D-glucosyl backbone of xyloglucan is synthesized in trans-cisternae, and the terminal fucosyl residues on the trisaccharide side chains of xyloglucan are partly added in the trans-cisternae, and partly in the trans-Golgi network (235). It has been shown by immuno-electron microscopy using anti- $\alpha\text{-L-Fuc-(1}\rightarrow 2)\text{-D-Gal}$ antibodies, that fucosylated XyG first appears in the lumen of the trans-Golgi and trans-Golgi network before vesicle mediated secretion

to the cell wall (224, 400). This activity appeared to be spatially distinct from galactosyl- and xylosyltransferase activity (401, 402).

5.6 Nucleotide sugars

The building blocks for polysaccharide synthesis are nucleotide sugars (NDP-sugars). The sugar moieties in NDP-sugars are incorporated into growing polysaccharide polymers by glycosyltransferases (GTs). A major contributor to glycan diversity is the number of NDP-sugars that an organism produces. In the plant kingdom, 30 different NDP-sugars have been identified (403). It is estimated that over 50 enzymes are directly involved in the synthesis of NDP-sugars in plants. To date only 22 NDP-sugar biosynthetic genes have been functionally identified [see (404) and text below] (Table 5.6). While it is widely accepted that NDP-sugars are the precursors for cell wall polysaccharides, glycoproteins, and glycolipids, it should be kept in mind that in addition to NDP-sugars, lipid-bound sugars are also sugar donors for the synthesis of glycans (for example, synthesis of the N-glycan core of glycoproteins in eukaryotes and addition of galacturonic acid to *Rhizobium* lipopolysaccharides via a prenyl phosphate–galacturonic acid donor substrate) (405, 406). Whether the initiation of plant cell wall polysaccharide synthesis is mediated by a core glycan that requires lipid-bound sugars at the ER, remains to be established.

Relatively few NDP-sugars are made inside the ER and Golgi apparatus where glycans are made. Most NDP-sugars are produced in the cytosol. Thus, specific NDP-sugar transporters exist to facilitate the import of NDP-sugars from the cytosol into the correct lumen of the endomembrane where GTs reside. It is predicted that ~20 NDP-sugar transporters exist in plants of which functionally only six have been characterized (A. Orellana, personal communication). These transporters are localized to the ER and Golgi apparatus (93, 407). While some of the NDP-sugar transporters are specific, the relatively low number of transporters would suggest that some NDP-sugar transporters may accept several NDP-sugars. In addition to the need for GTs, NDP-sugars, and their transporters for wall polysaccharide synthesis, some wall polysaccharides (i.e. pectins and hemicelluloses) are also modified by acetyl and methyl groups. Thus, diverse acetyltransferases and methyltransferases are also required. Little is known about their substrate specificity, as none has been biochemically purified. Basic questions such as what controls the degree and number of methyl modifications on a specific glycan remain elusive. Hence, no wall-related acetyl- or methyltransferases have been functionally cloned. The methyltransferases (MetTs) and acetyltransferases (AceTs) generally utilize *S*-adenosyl-*L*-methionine (SAM) and acetyl-CoA as methyl and acetyl donors, respectively (229, 285, 335, 337). A recent study in *A. Orellana's* laboratory led to the biochemical identification of a SAM transporter activity in the Golgi apparatus of pea (A. Orellana, personal communication). Beyond synthesis, plant glycans undergo further modification, including degradation and remodeling by specific glycosidases and esterases. Due to space limitations, transporters and glycan modifying enzymes will not be summarized in this review, rather, the reader is referred to a recent review (246).

Wall biogenesis is a complex cellular event similar to an assembly line. It requires the supply of a wide range of precursors targeted to different subcellular locations for a process that begins in one subcompartment and continues in other subcompartments as the glycans are synthesized and modified. During this process the concerted action of a range of cytosolic,

Table 5.6 NDP-sugar biosynthetic genes

Enzyme	Activity	Syn	Mutant, isoforms (putative?), locus	(aa)	Cell location
UDP-sugar PPase	Sugar-1-P + UTP ↔ UDP-sugar + PPi	Sloppy	Sloppy, At5g52560	614	
UDP-Glc PPase	Glc-1-P + UTP ↔ UDP-Glc + PPi	UGlc PPase	UGlcPP#1 At5g17310 UGlcPP#2 At3g03250	470 469	
Sucrose synthase	Sucrose + UDP ↔ UDP-Glc + Frc	SuSy	SuSy1, At5g20830 SuSy2, At3g43190 SuSy3, At5g49190 At4g02280	808 807 808 809	Mito Cyto Golgi
ADP-Glc PPase	Glc-1-P + ATP ↔ ADP-Glc + PPi	AGPase small sub AGPase large sub LS	ADG1, Aps1, At5g48300 Aps2, At1g05610? Apl1, ADG2 At5g19220 Apl2, At1g27680 Apl3, At4g39210 Apl4, At2g21590	520 476 521 523 521 520	Chlo Chlo Chlo
Galk	Gal + ATP ↔ Gal-1-P + ADP	Gal1, Galk	Galk, At3g06580	496	
UDP-Glc 4-epimerase	UDP-Glc ↔ UDP-Gal	UGE	UGE1, At1g12780 UGE2, At4g23920 UGE3, At1g63180 <i>Rhd1</i> , UGE4, At1g64440 UGE5, At1g10960	351 350 351 348 351	Cyto Cyto Cyto Cyto-Golgi associated Cyto
UDP-Rha synthase	UDP-Glc + NAD(P)H → UDP-Rha	URS Rhm	Urs1, At1g78570 At3g14790 <i>Rhm2</i> , <i>mum4</i> , At1g53500	669? 667	
UDP-Rha epimerase/reductase	UDP-4keto 6deoxyGlc + NAD(P)H → UDP-Rha	NRSer	NRS(er), At1g63000	301	
UDP-Glc dehydrogenase	UDP-Glc + 2NAD → UDP-GlcA + 2NADH	UGlcDH UGD	Ugd1, At5g39320 Ugd2, At3g29360? Ugd3, At5g15490 Ugd4, At1g26570	480 480 480 481	

UDP-GlcA 4-epimerase	UDP-GlcA \leftrightarrow UDP-GalA	UGlcAE GAE	UGlcAE1, At2g45310 UGlcAE2, Gae6, At3g23820 UGlcAE3, Gae1, At4g30440 UGlcAE4, At4g00110 UGlcAE5, Gae2 At1g02000? UGlcAE6, Gae5, At4g12250?	437 460 429 430 434 436	
UDP-GlcA decarboxylase	UDP-GlcA \rightarrow UDP-Xyl	UXS	Uxs1, At3g53520 Uxs2, At3g62830 Uxs4, At2g47650 Uxs3, At5g59290 Uxs5, At3g46440 Uxs6, At2g28760	433 445 443 342 341 343	Membrane Golgi Cytosol
UDP-Api/UDP-Xyl synthase	UDP-GlcA + NAD \rightarrow UDP-Api + UDP-Xyl	AXS UAS	AXS1, At1g08200 At2g27860	389 389	Cytosol
AraK	Ara + ATP \leftrightarrow Ara-1-P + ADP	Ara1	Ara1, At4g16130	1039	
UDP-Xyl 4-epimerase	UDP-D-Xyl \leftrightarrow UDP-D-Ara	UXE	Mur4, UXE1, At1g30620 UXE2, At2g34850? Uxe3, At3g34850 Uxe4, At5g44480?	419 379 385 436	Golgi
GDP-Man PCase	Man-1-P + GTP \leftrightarrow GDP-Man + PPi	GMPCase	Cyt1, At2g39770 At4g30570?	361 331	
GDP-Man 4,6-dehydratase	GDP-D-Man \rightarrow GDP-4keto 6deoxyMan	GMD	GMD1, At5g66280? GMD2, mur1, At3g51160	361 373	
GDP-Man 3,5 epimerase/ 4-reductase	GDP-4keto 6deoxyMan \rightarrow GDP-L-Fuc	GER	Ger1, At1g73250 Ger2, At1g17890	312 328	
GDP-Man 3',5' epimerase	GDP-D-Man \rightarrow GDP-L-Gal GDP-D-Man \rightarrow GDP-L-Gul	GME	Gme1, At5g28840	377	
KDO-8-P synthase	PEP + D-arabinose 5-phosphate \rightarrow KDO-8-P	kdsA	kdsA1, At1g79500 kdsA2, At5g09730	290 291	
CMP-KDO synthase	KDO + CTP \rightarrow CMP-KDO	kdsB	kdsB, At1g53000	290	

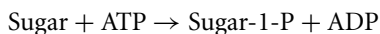
ER, and Golgi enzymes, as well as ER, Golgi, vesicular and plasma membrane proteins is required to facilitate the production of one type of glycan. Therefore, knowledge regarding the catalytic topology of each membrane enzyme and the cellular machinery that partitions, regulates, and traffics each protein and the corresponding glycan-intermediates, to their correct subcompartments must be understood to truly comprehend wall assembly and synthesis.

5.6.1 Fermentation and nucleotide-sugars: a long history

It was exactly 101 years ago, while working on fermentation of sugars by yeast, that Harden and Young (408) first reported the chemistry and metabolic roles of sugar-phosphates. Later, different types of sugar kinases from yeast, muscle, and plant sources were identified that were able to convert sugar (monosaccharide) and ATP to the phosphorylated-sugar esters. The seminal work by Cori *et al.* in 1939 (409) proved, for the first time, the role of sugar-phosphates in the synthesis of polysaccharides (i.e., glycogen). Since the discovery of UDP-glucose in yeast (410), which was followed by the isolation of other NDP-sugars in yeast, plants, bacteria, algae, and humans, it has become apparent that NDP-sugars are the prime sugar-substrates used in the biosynthesis of glycans. The biosynthesis of activated-sugars is achieved in three general ways.

5.6.2 Sugar kinase – pyrophosphorylase pathway to synthesize NDP-sugars

Some sugars are first converted directly, or in a series of enzymatic steps, to phospho-1-sugars in the presence of ATP.



Following the phosphorylation of the anomeric center, enzymes known as pyrophosphorylases transfer a nucleotide-monophosphate from NTP to the sugar-1-P to form the NDP-sugar. The synthesis of UDP-Glc and GDP-Man are examples of this type of synthesis.

5.6.3 Direct production of NDP-sugars

Some activated sugars such as CMP-KDO are exceptional since the free sugar, synthesized via intermediary metabolism products, is condensed directly with CTP without a prior phosphorylation of C1.



5.6.4 NDP-sugar Interconversion Pathway

The interconversion pathway of nucleotide-sugars is a major pathway where specific enzymes convert preexisting NDP-sugars into different stereospecific NDP-sugars (Figure 5.4).

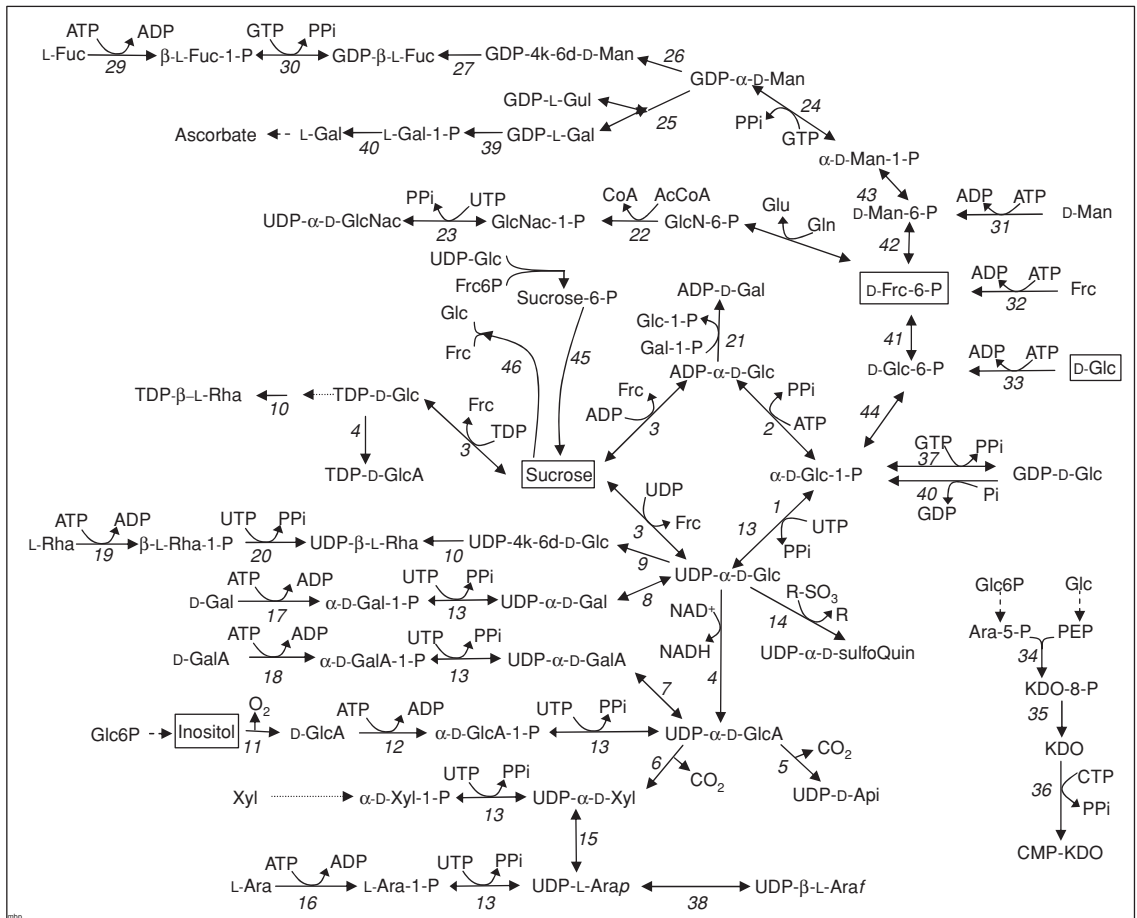


Figure 5.4 The metabolism of NDP-sugars in plants. Glucose, fructose, inositol, and sucrose are major sources of carbon that can be metabolized into NDP-sugars. A salvage pathway is defined as the ability of plant enzymes to recycle sugars that are released from glycoconjugates. The enzyme activity depicted as numbers (italic) are UDP-glucose pyrophosphorylase (1), ADP-glucose pyrophosphorylase (2), sucrose synthase (SuSy, 3), UDP-glucose dehydrogenase (UGD, 4), UDP-apiiose/xylose synthase (5), UDP-glucuronic acid decarboxylase (UXS, 6), UDP-glucuronic acid 4-epimerase (UGLAE, 7), UDP-galactose 4-epimerase (UGE, 8), UDP-glucose 4,6 dehydratase (9), UDP-4-keto-6-deoxyglucose-3',5'-epimerase and 4',6'-keto-reductase (NRSER, 10). Reaction 9 and reaction 10 are also carried out by a single polypeptide (URS/MUM/RHM, 9.10), inositol oxygenase (MIOX, 11), glucuronic acid-1-P kinase (12), UDP-sugar pyrophosphorylase (Sloppy, 13), UDP 5'-diphospho-6-sulfoquinovose synthase (14), UDP-xylose epimerase (UXE, 15), UDP-arabinomutase (UAM, RGP, 38), L-arabinose-1-P kinase (Ara1, AraK, 16), D-galactose-1-P kinase (GalK, 17), D-galacturonic acid-1-P kinase (GalAk, 18), L-rhamnose-1-P kinase (RhaK, 19), UDP-rhamnose pyrophosphorylase (20), AMP sugar-1-P transferase or ADP-glucose phosphorylase (21), glucoseamine-6-P acetyltransferase (GNA1, 22), UDP-N-acetylglucoseamine pyrophosphorylase (23), mannose-1-P pyrophosphorylase (24), GDP-Man 3',5' epimerase (GME, 25), GDP-L-galactose phosphorylase (39), L-galactose-1-P phosphatase (40), GDP-mannose 4,6 dehydratase (GMD, 26), GDP-4-keto-6-deoxymannose-3,5-epimerase-4-reductase (GER1, FX, 27), fucose-1-P kinase (29), GDP-fucose pyrophosphorylase (30), mannose-6-P kinase (31), fructose-6-P kinase (32), glucose-6-P kinase (hexokinase, 33), phosphoglucose isomerase (PGI, 41), phosphomannose isomerase (PMI, 42), phosphomannose mutase (PMM, 43), phosphoglucose mutase (PGM, 44), KDO-8-P synthase (34), KDO-8-P phosphatase (35), CMP-KDP synthase (36).

Some of the types of modifications involved in the interconversion pathway are the isomerization of L- and D-sugars, 4-epimerization, specific C-6 oxidation, decarboxylation, and the formation of NDP-deoxy sugar derivatives. Examples of the interconversion pathway are the conversion of UDP-GlcA to UDP-GalA and the conversion of UDP-GlcA to UDP-Xyl. The synthesis of each nucleotide-sugar is described in the following sections.

5.6.5 *SLOPPY, a general UDP-sugar pyrophosphorylase*

This section on NDP-sugar biosynthesis begins with a description of an enzyme we identified and characterized that does not obey the general rule of enzyme specificity (at least not as we were accustomed to it with other enzymes we had purified in the past). This enzyme, named SLOPPY, is responsible for the synthesis of at least six different UDP-sugars and is unique to the plant kingdom. No other genes with a high sequence similarity to Sloppy have been identified in other organisms to date.

D-Fructose-6-P, a product of photosynthesis, is a central precursor for all monosaccharide residues in plants (403). Using labeling experiments, however, it was also elegantly shown that plant cells can readily take up other free sugars such as Rha, Glc, Gal, Xyl, GalA, GlcA, Ara, and Fuc, and incorporate them into polysaccharides. It was assumed that a free sugar was first phosphorylated with ATP and then uridylated or guanylated with UTP or GTP to form the corresponding NDP-sugar. Indeed, numerous kinase and pyrophosphorylase activities were isolated in the late 1950s and early 1960s that catalyzed such reactions. The kinases were never purified and it was not explicitly clear if different kinases catalyzed the C-1 phosphorylation of each unique sugar or if one kinase phosphorylated several sugars. We discuss the different kinase activities and specificities below. Similarly, it was not clear if the subsequent pyrophosphorylation of each sugar-1-P by pyrophosphorylase was specific or not (411, 412). Three independent research groups identified an enzyme in pea (413) and in *Arabidopsis* (414, 415) that can pyrophosphorylate various sugar-1-phosphates with UTP. The recombinant *Arabidopsis* protein (At5g52560), termed SLOPPY, has a higher affinity (e.g., lower K_m) for GlcA-1-P but it also catalyzes the conversion of Glc-1-P, Gal-1-P, Xyl-1-P, Ara-1-P, and GalA-1-P to their respective UDP-sugars (414). The enzyme is very efficient and specific for the production of UDP-sugars and shows no detectable activity when TTP, GTP, ATP, CTP are substituted for UTP. Although Sloppy has broad sugar-1-P specificity, it cannot accept GalNac-1-P.

The existence of a non-specific pyrophosphorylase raises basic questions. What is the source of free sugars in plants? Do the free sugars contribute significantly to the flux of NDP-sugars for wall biosynthesis pathways? Are free sugars generated inside or outside the cells? If they are imported inside, are they derived from long-distances, cell-to-cell transport, or directly from recycled wall? These are central questions that both need to be addressed and obviously raise more questions. If indeed sugars are recycled from glycan degradation, are there sugar-specific transporters? Recently, a plasma membrane sugar transporter (AtPLT5) was functionally identified in *Arabidopsis* (416). The transporter is a member of a multigene family and seems to be a “non-specific transporter” since competition assays indicate that AtPLT5 can transport Xyl, Rib, Ara, Glc, and myo-inositol, but not sucrose. Unfortunately, the transport of other sugars such as Gal, Rha, and Fuc was not determined for AtPLT5. But it is likely that other sugars are transported either by this, or other, transporters.

The “recycling” of free sugars into the NDP-sugar pool was termed the “salvage pathway” (403). The relative amount of free sugars released from glycolipids, glycoproteins, wall polysaccharides, and glycosides of small metabolites is hard to quantify. Hence, at this time it is not possible to judge the relative contribution of the salvage pathway to the flux of NDP-sugars versus the carbon flux derived from photosynthesis. In the subsequent subsections we will describe the synthesis of specific NDP-sugars.

5.6.6 UDP- α -D-glucose (UDP-Glc)

UDP-Glc, the most abundant NDP-sugar, is the immediate precursor to UDP-Gal, UDP-Rha, and UDP-GlcA. UDP-Glc is produced via three separate metabolic routes. (i) UDP-Glc pyrophosphorylase (UGlCPP) catalyzes the formation of UDP-Glc and PPi from Glc-1-P and UTP in a reversible reaction. (ii) Sucrose synthase (SuSy) transfers the Glc moiety from sucrose onto a UDP acceptor, forming UDP-Glc and Fru. (iii) UDP-D-galactose-4 epimerase (UGE) reversibly epimerases UDP-Gal to UDP-Glc.

5.6.6.1 UDP-Glc Pyrophosphorylase (UGlCPP)

In *Arabidopsis*, two genes (At5g17310 and At3g03250) encode proteins that share high aa sequence identity to each other (93%) and to the well-characterized potato and barley UDP-Glc PPase (>80%). Recombinant At5g17310 (UGlCPP1) expressed in *E. coli* utilizes only Glc-1-P and UTP to form UDP-Glc. UGlCPP1 is specific for both UTP and Glc1P since TTP, GTP, ATP or other sugar-1-phosphates are not substrates for this enzyme (414). A crystal structure of UGlCPP2 and At3g03250 has been submitted (Wesenberg, G.E., Phillips, G.N., Jr., Bitto, E., Bingman, C.A., Allard, S.T.M.). Early biochemical work established that UDP-Glc PPase is inhibited by UDP-Xyl. If this inhibition occurs *in vivo*, it would imply that UDP-Xyl, in addition to gene expression, regulates the UDP-glucose pool, and thus, the NDP-sugar pool available for wall synthesis. Recent analysis of rice plants where one of the two rice UDP-Glc PPase genes, *ugp1*, was suppressed, suggests that the production of UDP-Glc during pollen development is critical for callose deposition (417).

5.6.6.2 Sucrose Synthase (SuSy)

Sucrose, a major carbon source for growing cells, is delivered between cells via plasmodesmata (symplastic route) and by long-range transport from source to sink cells via the phloem. SuSy, sucrose synthase (EC 2.4.1.13), catalyzes the reversible conversion of sucrose and UDP into UDP-Glc and fructose. But *in vitro* SuSy also converts sucrose to form TDP-Glc and ADP-Glc from TDP and ADP, respectively (403, 418), as well as GDP-Glc and CDP-Glc (419). SuSy isoforms have been identified in many plant species. In pea, three SuSy isoforms (Sus) were functionally isolated and found to have different kinetic properties. For example, Sus1 was strongly inhibited by Frc (420). In *Arabidopsis*, six distinct gene members of SuSy are known, and the tissue expression pattern for each SuSy transcript isoform is complex (421) and does not provide clue to their distinct biological functions. Part of the complexity in assigning a biological role for each isoform is the fact that SuSy isoforms differ in protein length and in their amino acid sequence. In addition, cell fractionation studies

and immunogold labeling demonstrate that SuSy isoforms are associated with different sub-compartments. For example, distinct SuSy isoforms fractionate with the Golgi apparatus, the tonoplast, and the plasma membrane (422). A recent report also identifies SuSy associated with the actin cytoskeleton (423). Subbaiah and coworkers (424) reported recently that of the three maize SuSy isoforms, SH1 contains a mitochondrial targeting peptide that is required for its intramitochondrial localization. This isoform was proposed to be involved in the regulation of solute fluxes into and out of mitochondria. The association of SuSy with membranes is often observed in growing cells. A specific phosphorylation site on the amino terminal region of SuSy regulates movement of the enzyme between a cytosolic form and a plasma membrane-associated form (425). Biochemical characterization and substrate specificities of each SuSy isoform will be required to elucidate their role in either supporting the flux of carbon to cell wall (i.e., A/T/UDP-Glc) or in carbohydrate storage (ADP-Glc). Lastly, an understanding of the relationship between UDP-Glc PPase isoforms and SuSy isoforms in photosynthetic and non-photosynthetic cells will be needed to understand how the flux of sugars is directed into growth or storage.

5.6.6.3 UDP-D-Galactose-4-epimerase (UGE)

The reversible UDP-Glc/UDP-Gal 4-epimerase (UGE) is a well-studied enzyme in all organisms. Enzyme characterization and crystal structures from numerous organisms revealed that UGE requires NAD^+ as a cofactor for the abstraction of the hydride from C4-OH to yield its 4-keto sugar ($\text{C}=\text{O}$) intermediate [for review see (426)]. Some UGEs require exogenous NAD^+ for activity while other types of UGE “tightly hold” the cofactor and require no supply of exogenous NAD^+ in the assay. After comparing UGE crystal structures from different organisms, Thoden and coworkers (427) suggested that the avidity of UGE for NAD^+ depends on the number of hydrogen (H) bonds each UGE has at the NAD -binding pocket. For example, the bacterial UGE “holds- NAD^+ ” with 19 H-bonds, whereas the human UGE which requires an exogenous supply of NAD^+ for activity, has only 11 H-bonds (427).

In plants, distinct UGE isoforms exist. UGEs were isolated from *Arabidopsis* (404, 428) and barley (426, 429) and multiple isoforms are found in the rice and maize genome. Recent thorough biochemical, genetic, and molecular studies led by G. Siefert on the five distinct UGE isoforms in *Arabidopsis* (404, 430) established that (i) isoforms differ in their requirement for exogenous NAD^+ , (ii) enzymatic efficiencies vary among the UGE isoforms, (iii) some isoforms likely exist to channel their enzymatic product for the synthesis of a particular glycan, (vi) transcripts of unique isoforms are expressed in the same cell suggesting that each is biologically distinct, and very interestingly (v) some isoforms, for example, UGE1, -2 and -4 are present in the cytoplasm, while UGE4 is enriched close to Golgi stacks. Barber and coworkers (430) suggest that plant UGE isoforms function in different metabolic situations and that enzymatic properties, gene expression patterns, and subcellular localizations contribute to the distinct isoform functions. This suggestion is likely correct, not only for UGE isoforms, but also for other NDP-sugar isoforms involved in the synthesis of specific plant NDP-sugars (431).

The root epidermal bulger 1 (reb1) mutant in *Arabidopsis thaliana* (432) is partially deficient in cell wall arabinogalactan-protein (AGP), indicating a role for REB1 in AGP biosynthesis (433). The REB1 is allelic to ROOT HAIR DEFICIENT 1 (RHD1) (434, 435), one of five ubiquitously expressed UGE genes. The RHD1 isoform is specifically required

for the galactosylation of xyloglucan (XG) and Type II arabinogalactan (AGII), but is not involved in either D-galactose detoxification or in galactolipid biosynthesis. Epidermal cell walls in the root expansion zone lack arabinosylated 1,6-linked- β -D-galactan and galactosylated XG. In cortical cells of *rhd1*, galactosylated XG is absent, but an arabinosylated 1,6-linked- β -D-galactan is present. These results show that the flux of galactose from UDP-D-Gal into different downstream products is compartmentalized at the level of cytosolic UGE isoforms and suggest that substrate channeling plays a role in the regulation of plant cell wall biosynthesis (435).

5.6.7 ADP- α -D-glucose (ADP-Glc)

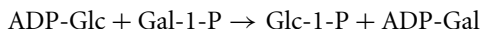
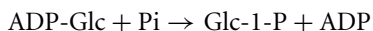
ADP-Glc is the major precursor for starch synthesis. Starch is a polymer of α -D-Glc consisting of two types of molecules: amylose (linear α -1,4-linked glucose) and amylopectin in which one Glc in every 20 or so residues on an amylose-like structure is branched by an α (1-6)-linkage connected to an α -1,4-linked chain. Although starch synthesis is not related to pectin synthesis, the information available about the regulation of the synthesis of the starch precursor, ADP-Glc, can be used as a paradigm when considering the synthesis and regulation of other NDP-sugars.

Starch, the main carbon storage form in plants, is made in plastids of photosynthetic and non-photosynthetic tissues. Adenosine 5'-diphosphate:glucose pyrophosphorylase (ADPGlc:PPase) catalyzes the first and rate limiting step in starch biosynthesis (i.e., the conversion of Glc-1-P and ATP to ADP-Glc and pyrophosphate (436). In cereal endosperms, two distinct ADPGlc PPases exist, one is found in the cytosol and the other in plastids (437–439). By contrast, ADPGlc PPase is exclusively located in plastids of leaves of both mono- and dicotyledonous plants, as well as in heterotrophic organs of dicotyledonous plants. Plant ADPGlc PPase is composed of two types of subunits (small and large) and is allosterically regulated by 3-phosphoglycerate and phosphate. The Arabidopsis genome consists of six ADPGlc PPase-encoding genes (two small subunits, ApS1 and ApS2; and four large subunits, ApL1–ApL4). Based on recombinant enzyme activities, mRNA expression and the fact that recombinant ApS2 has no ADPGlc PPase activity, it has been proposed (440, 441) that ApS1 is the main catalytic isoform responsible for ADPGlc PPase activity in all tissues of Arabidopsis. The authors suggested that each isoform of the large subunits plays a regulatory role. The large subunit, ApL1 is expressed in source tissues, whereas ApL3 and ApL4 are the main isoforms expressed in sink tissues. Thus, in source tissues, ADPGlc PPase could be regulated by the 3-phosphoglycerate/phosphate ratios, while in sink tissues; the enzyme would be dependent on the availability of substrates for starch synthesis.

In cereal endosperm, on the other hand, a different regulation of starch synthesis may operate. It appears that the transport of ADP-Glc from the cytosol into the plastid is the limiting factor. This became clear during the characterization of a plastidial ADPGlc transporter (HvNst1) barley mutant with low-starch content (442). The mutant accumulates high levels of ADP-Glc in the developing endosperm indicating that cytosolic pool of ADP-Glc is not under metabolic control in this tissue. Lastly, leaves overexpressing SuSy showed a large increase in the levels of both ADP-Glc and starch, compared with WT leaves, while leaves overexpressing antisense SuSy accumulated low amounts of both ADP-Glc and starch (438). The above findings, which originated in the Pozueta-Romero's laboratory, show that in source leaves ADP-Glc produced by SuSy (outside the chloroplast) is directly linked to,

and appears to control starch biosynthesis. This implies that SuSy, but not ADPGlc:PPase, controls the level of ADP-Glc in the cytosol in source leaves (438, 439).

More recently, a new enzyme activity was identified in *Arabidopsis* that must be considered to better evaluate the metabolic fate of ADP-Glc in the cytosol. Recombinant *Arabidopsis* At5g18200 has ADP-Glc phosphorylase activity (please note it is not a PPase). The enzyme is capable of transferring AMP from ADP-Glc onto either Pi or Gal-1-P (443) as shown in the scheme below:



Unlike the human and fungal GalT enzyme, which transfers UMP from UDP-Glc onto Gal-1-P forming Glc-1-P and UDP-Gal, the ADP-Glc phosphorylase cannot utilize UDP-Glc as a donor substrate (443).

We put forward that the above-described distinct regulatory role of ADPGlc PPase and SuSy are examples that highlight the possibility that different isoforms of nucleotide-sugar biosynthetic enzymes may have distinct roles in plants and that different plant species may regulate the same metabolic pathway in different ways.

5.6.8 UDP- α -D-galactose (UDP-Gal)

Galactose is a major constituent of diverse pectic polysaccharides including RG-I. The sugar donor, UDP-Gal, is produced by (i) phosphorylation of C-1 of D-Gal by galactokinase (414) followed by pyrophosphorylase-catalyzed conversion of α -D-Gal-1-P and UTP to UDP-Gal, and by (ii) UDP-Glc-4 epimerase (UGE as described above) that reversibly converts UDP-Glc to UDP-Gal.

- 1 D-galactokinase activity was isolated by Neufeld and coworkers in 1960 from mung bean (412). The D-GalK is membrane bound and the activity, unlike L-AraK activity, could not be solubilized with digitonin. The galactokinase gene (GalK, Gal1, At3g06580) was cloned from *Arabidopsis* by functional complementation of yeast (446) and *E. coli* (445) galK mutants that are unable to metabolize galactose. While the Gal1/GalK clone was able to complement the yeast mutant, a definitive substrate specificity of the recombinant *Arabidopsis* enzyme (GalK) will provide information on whether other “recycled sugars” are substrates. Sequence alignment of various sugar kinase proteins shows that the *Arabidopsis* GalK shares amino acid sequence similarity (45%) to GalK2, a human kinase with phosphorylation preference to GalNAc (446). A meaningful alignment could not be obtained between the *Arabidopsis* GalK with the human GalK1 whose true substrate is Gal (447). Whether At3g06580 encodes a Gal-1-P kinase activity and/or GalNAc-1-P kinase activity remains to be determined biochemically. The subsequent pyrophosphorylation of Gal-1-P to UDP-Gal is likely mediated via “Sloppy (416),” the non-specific UDP-sugar pyrophosphorylase (At5g52560).
- 2 The second route to form UDP-Gal is with UDP-Glc-4-epimerase (as described above). In humans and fungi, UDP-Gal is synthesized by uridylyltransferase activity (GalT). GalT transfers UMP from UDP-Glc onto Gal-1P forming Glc-1-P and UDP-Gal. Such activity and corresponding genes have not yet been described in plants.

5.6.9 UDP-L-rhamnose (UDP-Rha)

Rhamnose is a major sugar moiety in pectin and in various glycosides of secondary metabolites. UDP-Rha is the activated sugar for the synthesis of flavonoids (448); however, the activated rhamnose-donor form for pectin synthesis has not been determined. Previously, it was suggested that synthesis of UDP-Rha from UDP-Glc is mediated by three separate enzymes, similar to the conversion of TDP-Glc to TDP-Rha in bacteria (403). UDP-Glc is first modified to the UDP-4-keto-6-deoxyGlc intermediate by UDP-Glc 4,6-dehydratase. The intermediate is modified in the presence of NAD(P)H by a 3,5-epimerase and 4,6-keto-reductase to form UDP-L- β -Rha. A debate in the literature as to whether two or three different enzymes are involved in UDP-Rha came to an end with the functional cloning NRS/er (134) from Arabidopsis (At1g6300). The activity of recombinant NRS/er demonstrates irrefutably that the 3,5-epimerase and 4,6-keto-reductase activities reside in one polypeptide. Interestingly, *in vitro* NRS/er accepts both TDP- and UDP-4-keto-6-deoxyGlc as substrates to form TDP-Rha and UDP-Rha, respectively. Although TDP-Glc is found in plants (403) and several enzymes can generate TDP-Glc *in vitro*, the physiological significance of the ability of NRS/er to generate TDP-Rha is unclear. Only by isolating a pectin rhamnosyltransferase and characterizing the donor specificity, can the true nature of NDP-Rha form be conclusively determined.

The Arabidopsis genome consists of three genes (At1g78570, At3g14790, At1g53500) each that encodes a large protein (~670 aa) having two domains: an N-terminal domain (~330 aa long) that shares amino acid sequence similarity to 4,6-dehydratase followed by a C-terminal domain (~320 aa long) that shares over 80% sequence identity to NRS/ER. The C-terminal domain of At1g78570 has similar enzyme activity as NRS/ER (134). Mutations in At1g53500, *mum4* (449), and *rhm2* (450) result in decreasing amounts of Rha and GalA sugar moieties in RG-I structures isolated from seed mucilage. These mutants provided the first genetic evidence for the involvement of these genes (we named URS, UDP-Rha-synthase) in rhamnose synthesis. More recently, when these genes were recombinantly expressed in yeast, Oka and coworkers (451) reported that all of the Arabidopsis URS genes (also named RHM/MUM) have UDP-rhamnose synthase activity and interestingly, are highly inhibited by UDP-Xylose.

5.6.10 UDP- α -D-glucuronic acid (UDP-GlcA)

In 1952, Dutton and Storey discovered that UDP-GlcA acts as a glucuronosyl donor in the synthesis of glucuronides by liver enzymes. In plants, UDP-GlcA is a key intermediate serving as a branch point between UDP-hexose (six carbons) and UDP-pentose (five carbons) sugars. UDP-GlcA is the precursor for UDP-D-xylose, UDP-L-arabinose, UDP-apiose, and UDP-galacturonic acid that contribute to synthesis of over 40% of cell wall polysaccharides. UDP-GlcA is made by (i) sequential phosphorylation of D-GlcA at C-1 by a membrane bound kinase (411) followed by a pyrophosphorylase activity that converts α -D-GlcA-1-P and UTP to UDP-GlcA, (ii) NAD-dependent oxidation of UDP-Glc to UDP-GlcA by UDP-Glc dehydrogenase (UGD, UDPGDH), and by a controversial pathway (iii) oxidation of UDP-Glc to UDP-GlcA by a bifunctional alcohol dehydrogenase ADH/UDPGDH.

In the 1960s the pathway from myo-inositol to cell wall glycans was proposed as a significant metabolic pathway. Early experiments with ^3H -inositol demonstrated that the label

was readily incorporated into cell wall polysaccharides (452). More recently, a labeling experiment with inositol in *Arabidopsis* showed that radioactivity is found only in the uronic acids, arabinose, and xylose that were released from wall glycans (453). The ability of inositol to drive the synthesis of GlcA was named “the myo-inositol oxidation pathway.” We will discuss the myo-inositol and salvage pathways separately.

5.6.10.1 *The myo-inositol pathway*

In plants, the first step in myo-inositol synthesis is the cyclization of D-Glc-6-P to myo-inositol-1-P (Ino-1P) by 1L-myo-inositol 1-phosphate synthase. In *Arabidopsis*, two functional isoforms were reported, At4g39800 and At2g22240 (454). The second step involves dephosphorylation of Ino-1-P to myo-inositol by myo-inositol monophosphatase (IMPase; EC 3.1.3.25) (455). Distinct multiple but highly conserved IMPase isoforms are found in each plant species. Three IMPases were identified in tomato (456). In *Arabidopsis*, a conserved IMPase-like-protein (At3g02870) was proposed by Glilaspi to act as IMPase; however, biochemical and genetic data indicate that At3g02870 encodes L-Gal-1-P phosphatase (457, 458). It is possible that other IMPase-like proteins in *Arabidopsis* (for example, At1g31190, At4g39120) encode the Ino-1P phosphatase activity to form myo-inositol. The identification of the true IMPase gene product is critical to evaluate what controls the pathway to shunt Ino-1P to the myo-inositol oxidation pathway. Free myo-inositol is oxidized by inositol oxygenase (MIOX; E.C. 1.13.99.1) to D-GlcA. *Arabidopsis* contains four Miox isoforms (453).

It would be interesting to determine if the myo-inositol oxidation pathway operates independently of the pathway leading to synthesis of UDP-GlcA from UDP-Glc. This knowledge could aid in determining which flux of sugars the plant uses to facilitate wall synthesis in specific tissues. For example, myo-inositol in seed is stored as phytic acid (inositol hexaphosphate). During germination, phosphatases provide a rapid source of inositol which is converted, in part, to GlcA. Hence, this would provide a source of UDP-GlcA for wall pectin synthesis. However, during germination, rapid synthesis of L-ascorbate from myo-inositol also occurs. The relationship and coordination of the supply of sugars to wall glycans and to ascorbate synthesis must be better understood at all stages of growth.

5.6.10.2 *The salvage pathway*

Feeding experiments with free D-GlcA showed that plant cells rapidly incorporate the sugar into pectin (459). Free GlcA, likely released from wall polysaccharides, can be activated to UDP-GlcA, by-passing the need for flux of the sugar from the inositol oxidation pathway. Membrane and soluble protein preparations extracted from a 4-day-old etiolated seedlings of mung bean consist of a kinase activity that specifically phosphorylates GlcA in the presence of ATP and Mg^{2+} to GlcA-1-P. L-Ara and D-Gal are also substrates for this crude kinase preparation. However, GalA is not a substrate for this activity (411). The GlcA-1-P kinase has not been purified and the gene is not yet known. Work in our laboratory (Bar-Peled, 2005) (414) and by Schnurr and coworkers (415) demonstrated that Sloppy (At5g52560) is the true GlcA-1-P pyrophosphorylase and readily converts UTP and GlcA-1-P into UDP-GlcA.

5.6.10.3 *UDP-Glc dehydrogenase (UGD)*

UDP-Glc dehydrogenase (UDPGDH, UGD) catalyzes the C-6 oxidation of UDP-Glc in the presence of two molecules of NAD^+ to UDP-GlcA and 2NADH. The enzyme is well

characterized from numerous organisms and crystal structures are available. In plants, the gene encoding this activity was first described in soybean (460), and soon after, as genome sequence became available, numerous UDPGDH gene isoforms that share high amino acid sequence similarity to each other were identified in Arabidopsis (461), poplar (462), tobacco (463), maize (464), and Dunaliella (465). In Arabidopsis, four UDPGDH isozymes are present and, likewise, multiple UDPGDH isoforms are found in the genome of every plant species that has been sequenced. The role of multiple UDPGDH isoforms and their specific activities is of great interest since recent work in sugar cane revealed that CTP-glucose and TDP-glucose were oxidized as well, although at low rates compared to UDP-Glc (466). The current knowledge indicates that UDPGDH is cytosolic and its activity is strongly inhibited by a very low level of UDP-Xyl. This suggests that flux of UDP-Glc (hexose) to UDP-pentoses can be feedback-regulated, in part, at the enzyme level.

The source of UDP-GlcA (i.e., from myo-inositol or UDP-Glc) for wall synthesis was elegantly determined by a genetic approach. Two UDP-Glc dehydrogenase mutants were identified in maize, one in isoform A and the other in isoform B of UDPGDH. Polysaccharides isolated from isoform A mutant had lower Ara/Gal and Xyl/Gal ratios when compared with wildtype, indicating the importance of UDPGDH in directly providing major flux of NDP-sugars to wall polysaccharides (464). On the other hand, the lack in "wall alteration" in the isoform B mutant, could suggest either that different UDPGDH isoforms function in a different metabolic pathways, for example in the formation of glycosides, or that isoform B contributes to the synthesis of a glycan structure that is present at such low levels that small alterations in the amount of GlcA was undetected. RG-I, for example, consists of a small amount of GlcA residue (only 1% of total sugars). Similarly, GlcA residues are found in relatively low amounts in RG-II and xylan, representing 3 and 5%, respectively, of total sugars in those polysaccharides.

In 1996, Robertson and coworkers (467) reported that alcohol dehydrogenase (ADH) from *Phaseolus vulgaris* can convert UDP-Glc to UDP-GlcA. This bifunctional activity of ADH to oxidize both ethanol and UDP-Glc opened a debate in the literature. A major difficulty in interpreting this result is that Robertson and coworkers (467) used a spectrophotometer assay which measures NADH formation, not UDP-GlcA directly. Recently, the tobacco ADH homologous genes, cloned and expressed in bacteria, had activity on both ethanol and UDP-Glc. But again, the UDPGDH assay was determined by spectrophotometer assays (463). Whether recombinant ADH catalyzed the formation, UDP-GlcA was not explicitly confirmed. Further kinetic work with pure ADH must be carried out to claim the dual specificity of this enzyme. Only if knockouts of all ADH genes are obtained can it be definitely determined whether ADH is a bifunctional enzyme. Maize mutants lacking ADH1 and ADH2 isoforms, however, had no effect on sugar composition of hemicellulose (464). If only two ADH isoforms exist in maize ADH, this would suggest that ADH is not a bifunctional enzyme.

5.6.10.4 Summary regarding UDP-GlcA formation

What controls the supply and flux of UDP-GlcA in plants is still debatable and it is very likely that different plant species adopt different mechanisms to control the supply of UDP-GlcA. In maize, mutants lacking the activity of one UDPGDH isoform have a reduction in the content of Ara, and Xyl in hemicellulose. This suggests that in maize, UDPGDH-A is a major supplier of the UDP-pentoses and that the myo-inositol oxidation pathway (not

“bifunctional ADH”) cannot compensate the flux of “sugar” to the formation of UDP-GlcA. In *Arabidopsis* mutants lacking MIOX1 and 2 isoform activities, no significant differences in monosaccharide amount or composition were observed in wall polysaccharides when compared to wildtype. Thus, it is likely that the major contributor for flux of NDP-sugars in plants is UDP-Glc. As mentioned above, it is possible that the myo-inositol pathway and the salvage pathway operate in a tissue-specific manner, for example, during pollen tube growth. During pollen tube germination and growth, large amounts of pectin are degraded. It is likely that the free sugars are recycled back by kinases and the activity of Sloppy, to readily form an available pool of NDP-sugars. This pool will provide NDP-sugars for growth of the pollen tube. Pollen tubes are one of the fastest growing cells known (1 cm h^{-1}) and mutants lacking Sloppy have a pollen phenotype (415). A similar regulation scenario likely occurs during seed germination with seeds that store a large amount of phytic acid, as discussed above.

5.6.11 UDP- α -D-galacturonic acid (UDP-GalA)

UDP-GalA, a major sugar donor for pectin synthesis, is made by (i) the salvage pathway, by the phosphorylation of D-GalA to GalA-1-P in the presence of ATP and kinase activity (GalAK). The GalA-1-P pyrophosphorylase catalyzes the conversion of GalA1P and UTP to UDP-GalA. (ii) UDP-GlcA 4-epimerase (UGlCAE), a reversible 4-epimerase that converts UDP-GlcA to UDP-GalA.

- 1 Feeding experiments with radioactive GalA demonstrated that the label is readily incorporated into pectin (459). Soluble enzyme preparations from a 1-day-old germinating seeds of mung bean have high GalAK activity (468), while no GalAK is observed in a 4-day-old etiolated seedlings (411). The collective biochemical data suggest that the soluble GalA kinase differs from the membrane-associated kinase activity that phosphorylates GlcA. A functional gene encoding GalAK activity was recently discovered in *Arabidopsis* [Ting and Bar-Peled, 2006, unpublished]. Proton NMR analysis confirm that pure recombinant GalAK phosphorylates D-GalA in the presence of ATP to GalA-1-P. Currently however, it is difficult to predict the relative amount of GalA that is recycled back to UDP-GalA. Clearly, high activities of pectin degrading enzymes during pollen germination could provide sufficient substrate for GalAK. Indeed, the non-specific UDP-sugar pyrophosphorylase (fondly named SLOPPY in *Arabidopsis* (414) and also identified in pea (469)) that converts GalA-1-P with UTP to UDP-GalA is highly expressed during pollen germination. The relative amount of GalA recycled from pectins as free sugars in other plant tissues is unknown, although the “fate” of free GalA could be tissue specific. For example, during strawberry fruit maturation, an increase in GalA-reductase (GalUR) activity was shown to direct GalA, released from pectin, for the synthesis of ascorbic acid (470). Thus, the relative contribution of the GalA-salvage pathway for synthesis of wall polymers remains unclear. Hopefully, current analyses of GalAK mutants in *Arabidopsis* will shed light on its physiological function (i.e. in the wall and in ascorbic acid metabolism).
- 2 UDP-GlcA 4-epimerase (UGlCAE) is a membrane-bound enzyme that reversibly catalyzes the 4-epimerization of UDP-GlcA to UDP-GalA. In *Arabidopsis*, six distinct functional genes (UGlCAe) encode isoforms having UDP-GlcA 4-epimerase activity (450, 471, 472). The isoforms can be divided into three evolutionary clades: Type A, B, and C. Members of UGlCAE are predicted to be Type II membrane proteins, suggesting that their catalytic

domain faces the lumen of an endomembrane. The topology of the catalytic domain was not validated experimentally. Mohnen's laboratory has shown that the activity of UGlcAE co-fractionated with the Golgi on sucrose-gradients (230) and expressing UGlcAE1-GFP in plants demonstrates that the chimeric fusion targets UGlcAE1 to the Golgi as well (Gu and Bar-Peled, unpublished data). Multiple UGlcAE isoforms are common and found in other plants, for example, rice and maize. Biochemical analyses of Arabidopsis, rice, and maize UGlcAE isoforms showed that the reversible UGlcAE epimerase has a preference (2:1) to form UDP-GalA and is inhibited by both UDP-Ara and UDP-Xyl. Recent studies (Gu and coworkers, submitted) demonstrated that the maize UGlcAE is especially sensitive to, and strongly inhibited by, UDP-Xyl when compared with Arabidopsis UGlcAE2. The authors suggest that in maize, the relative low amount of pectin (when compared to dicots) could be due to inhibition of UGlcAE by UDP-Xyl. Hence, this may provide an explanation as to why maize has less pectin and more xylan when compared to dicots. The role of multiple UGlcAE isoforms is currently being addressed. One possible explanation is that different isoforms localize to distinct endomembranes, as was suggested by Pattathil and coworkers (473) for the different Uxs isoforms.

5.6.12 UDP- α -D-xylose (UDP-Xyl)

UDP-Xyl is primarily synthesized by UDP-GlcA decarboxylase (UGlcA-DC, also named UDP-Xylose Synthase, Uxs) from UDP-GlcA. The enzyme has a tightly bound NAD^+ , which participates first in the oxidation of UDP-GlcA to the UDP-4ketohexose intermediate resulting in decarboxylation (removal of COOH as gas, CO_2) and formation of UDP-4ketopentose. In the second stage, the NADH-bound enzyme reduces the UDP-4ketopentose to UDP-Xyl resulting in the release of NAD-bound enzyme (403). Multiple distinct UXS isoforms encoding this enzyme activity were reported in Arabidopsis (431), rice (474), barley (475), and tobacco (463). Uxs isoforms are very specific enzymes and act only on UDP-GlcA. The 4-epimer of UDP-GlcA, UDP-GalA is not a substrate for Uxs (431). Uxs is active as a dimer and inhibited by UDP-Xyl. In plants, UDP-Xyl is made in the cytosol and in the endomembrane system (473). Phylogeny analysis classified the six Uxs isoforms from Arabidopsis into three distinct clades: Type A (1 isoform, At3g53520); Type B (2 isoforms, At3g62830, At2g47650), and Type C (3 isoforms, at5g59290, At3g46440, At2g28760). Type A and B isoforms have an N-terminal extension (~ 120 aa long) which encodes longer proteins compared to Type C Uxs isoforms. Type A and B UXS isoforms are predicted Type II membrane proteins with the catalytic domain facing the endomembrane lumen (431). Expression of Type B Uxs isoforms in plants confirmed that Uxs2 is an integral membrane protein. Expressing of a Uxs2-GFP construct in tobacco leaves was shown to localize the chimeric fusion protein in the Golgi apparatus (473). Uxs3 belongs to the Type C clade and the isoform was found in the cytosol as predicted.

5.6.13 UDP-D-apirose (UDP-API)

UDP-API Synthase (UAS, AXS) converts UDP-GlcA in the presence of NAD^+ to UDP-API. The enzyme decarboxylates UDP-GlcA to form a UDP-4ketopentose intermediate and the release of CO_2 , and then catalyzes rearrangement of the sugar skeleton to form UDP-API

(476, 477). In vitro, the enzyme forms both UDP-Xyl and UDP-Api (403). We believe that the formation of UDP-Api was not confirmed satisfactorily since the product is readily degraded to cyclic apiose 1,2-phosphate. Functional genes encoding AXS were isolated from tobacco (358), Arabidopsis and potato (Guyet and Bar-Peled, unpublished) (478). Two isoforms exist in Arabidopsis (At1g08200 and At2g27860) and are predicted to be in the cytosol. The “dual function” of the enzyme in generating both UDP-Xyl and UDP-Api was assayed by NMR spectroscopy (Guyet and Bar-Peled, unpublished). NMR time course assays for the conversion of UDP-GlcA into UDP-pentose using recombinant potato UAS, confirms explicitly that UDP-Api is made first. The analysis indicates that in vitro, UDP-Xyl synthesis lags behind UDP-Api. Mutation in UDP-Api synthase in *Nicotiana benthamiana* is lethal as a consequence of the lack of RG-II (358). To unambiguously determine if the enzyme is bifunctional and contributes to UDP-Xyl synthesis, a mutation in the three cytosolic UXS (Type C) genes must be carried out.

5.6.14 UDP-L-arabinose pyranose (UDP-Ara)

Arabinose (Ara) is an important sugar in plant walls and with a few exceptions, the predominant form of Ara in plant glycans is the furanose configuration (Ara). However, some polysaccharides, RG-II, for example, carry both forms of the arabinose moiety, i.e. Ara-furanose and Ara-pyranose. UDP-Ara (pyranose form) was identified in all plant extracts and is synthesized by (i) the sequential phosphorylation of L-Ara at its C-1 by a membrane associated L-arabinokinase (412) followed by a pyrophosphorylase that converts L-Ara-1-P and UTP to UDP-Arap, and (ii) a membrane-bound UDP-Xyl-4 epimerase (UXE) that converts UDP-D-Xyl to UDP-L-Arap.

1 Neufeld and coworkers (411, 412, 468) isolated sugar-kinase activities from different sources of tissues. A membrane fraction from bean was shown to catalyze the C-1 phosphorylation of L-Ara to β -L-Ara-1-P. The same membrane preparation phosphorylated D-Gal to α -D-Gal-1-P. However, the GalK kinase and the AraK kinase are different enzymes since AraK requires divalent ion for activity (Mg^{2+} , Mn^{2+}) whereas, the GalK kinase requires no additional divalent ion for activity (412). In addition, treatment of membranes with digitonin solubilizes the L-AraK activity but not the D-GalK activity. The AraK is specific for the L-form since D-Ara (that is common in prokaryotes) is not a substrate.

The *ara1* mutant from Arabidopsis, in the At4g16130 locus, has reduced ability to metabolize arabinose and lacks Ara-1-P kinase activity (479). Bioinformatic analysis suggests that Ara1 belongs to a large family including galactokinase, homoserine kinase, mevalonate kinase, and phosphomevalonate kinase (GHMP kinases). The Ara1 protein is speculated to be a Type Ia membrane protein. If the topology is correct, it would be interesting to know whether the catalytic domain is facing the cytosol or the lumen. Direct biochemical assays and substrate specificity studies of the encoded Ara1 gene were not performed. The subsequent pyrophosphorylation of α -L-Ara-1-P to UDP-Ara can be mediated by “Sloppy,” the non-specific UDP-sugar pyrophosphorylase (413).

2 A genetic screen led by Reiter *et al.* (125) identified the *mur4* mutant in Arabidopsis that has a 50% reduction of arabinose in the wall. The encoded recombinant protein (Uxe1,

UDP-Xyl 4-epimerase) was shown to convert reversibly UDP-D-Xyl to UDP-Ara (480). Bioinformatic analysis suggests that Uxe1 is a Type II membrane protein whose catalytic domain is facing the lumen. A Uxe1-GFP chimera was localized to the Golgi apparatus (480). Two isoforms (UXE1, At1g30620 and UXE2, At2g34850) that share 83% aa sequence identity to each other exist in the Arabidopsis genome; two isoforms are in the rice genome and three UXE isoforms were isolated for barley (*Hordeum vulgare*) (Zhang and Fincher, unpublished). Since several GTs were able to transfer the Ara (pyranose) from UDP-Ara into plant glycans (271, 397); it remains a puzzle when the Ara acquires the furanose configuration. One can predict that the Araf-donor has not yet been identified. However, this is unlikely since the *mur4* mutant (involved in the synthesis of UDP-Ara pyranose), lacks glycan consisting of Araf. This could imply that during the arabinosyltransferase catalyzed reaction the Araf is altered to the Araf form. A specific mutase may exist to convert the Araf to Araf on the glycan itself, similar to the conversion of GlcA to IdoA in proteoglycans (481). Alternatively, UDP-Ara (furanose) is made in plants as recently confirmed and described in the following section.

5.6.15 UDP-arabinose furanose (UDP-Araf)

Recent work in Tadashi Ishi's laboratory established a new enzyme activity that was never reported before, a UDP-arabinopyranose mutase (UAM). The enzyme is capable of converting UDP-Ara pyranose to UDP-Ara furanose (UDP-Ara *f*) and it was reported to be a reversible reaction (393). This activity was biochemically purified from rice and the corresponding gene was cloned. In rice, two homologous proteins (UAM1; AK098933, UAM2; AK071012) were identified. The recombinant enzyme at thermodynamic equilibrium produces UDP-Ara in a pyranose:furanose ratio of 90:10. The Arabidopsis homologous proteins were initially named as reversibly glycosylated protein, RGP (396, 484). Several isoforms of RGPs exist in the plant kingdom. The Arabidopsis RGP1 (At3g02230) and RGP2 (At5g15650) proteins were found to localize in the Golgi apparatus (396, 483). However, Sagi and coworkers (484) observed that a chimeric RGP-tagged to green fluorescence protein is localized to both Golgi and plasmodesmata. The specific *in vivo* role of RGP/UAM and the various isoforms in the (i) synthesis of the furanose form of UDP-Ara or as (ii) a reversibly glycosylated protein remains unclear.

5.6.16 GDP- α -D-mannose (GDP-Man)

GDP-Man (485) is a major sugar donor that provides flux of sugar to the synthesis of glycoproteins, polysaccharides, and ascorbic acid in plants. GDP-Man is the precursor for GDP-Fuc and GDP-Gal. Guanosine 5'-diphosphate-mannose pyrophosphorylase (GDP-Man PPase) catalyzes the conversion of α -D-Man-1-P and GTP to GDP-Man and pyrophosphate (486). The enzyme activity requires Mg^{2+} and the enzyme appears to be cytosolic. The pyrophosphorylase is well studied in various organisms and the crystal structure is known. In Arabidopsis, a wall mutant *cyt1* (487), an ozone-sensitive mutant, and ascorbic acid mutant, *vtc1*, identified the same gene product At2g39770 as the locus responsible for the production of GDP-Man (488, 489). The mutant likely survives since another homologous gene, At4g30570, may compensate for its activity. Although the enzymatic activity

of recombinant protein was not described, work in our laboratory confirmed that the encoded gene At2g39770 has GDP-Man PPase activity and is very specific toward Man-1-P as a substrate (Echale and Bar-Peled, unpublished).

5.6.17 GDP- β -L-fucose (GDP-Fuc)

Plant cell wall polysaccharides contain L-fucose (6-deoxy-L-Gal) derived from the sugar-donor, GDP-Fuc. GDP-Fuc synthesis occurs in two enzymatic steps (similar to the synthesis of UDP-Rha from UDP-Glc). These enzymes have been characterized and the corresponding functional genes have been identified in humans, plants, and bacteria. First, GDP-Man 4,6-dehydratase (GMD), converts GDP-Man to a GDP-4keto-6-deoxyMan intermediate. The latter is then converted by GDP-4-keto-6-deoxymannose-3,5-epimerase-4-reductase (GER1, FX) to GDP-Fuc. In Arabidopsis, there are two GMD isoforms (GMD1, At5g66280; GMD2 (*mur1*) At3g51160) which share 92% aa sequence identity to each other; and two GER isoforms (GER1, At1g73250; GER2, At1g17890) that share high (88%) sequence identity to each other. In some tissues it appeared that GMD isoforms are co-expressed, but in other tissues expression is restricted. For example, GMD2 is expressed in most cell types of the root, but not in the root tip where strong expression of GMD1 is observed (490). Within shoot organs, GMD2 appears to be expressed in most tissues while GMD1 expression is restricted to stipules and pollen grains. The lack of GMD2 above ground (*mur1* mutant) corresponds to an almost complete reduction in Fuc in wall polysaccharides including XG whose Fuc can be substituted by L-Gal presumably as a result of increased GDP-L-Gal availability (491, 492). However, below ground the *mur1* mutation leads to a 40% reduction of Fuc. Some isoforms may have redundant function in a specific cell, but in other tissues of the same plant the existence of isoforms may provide pools of the NDP-sugars to synthesis of different types of glycans.

5.6.18 GDP- β -L-galactose (GDP-Gal), GDP- β -L-glucose gulose (GDP-Gul)

GDP-Gal is a major precursor for the synthesis of ascorbic acid in plants, and relatively low amounts of L-Gal are found in plant glycans. A GDP-Man 3',5' epimerase activity first identified in the Neufeld's laboratory (493) epimerizes GDP-Man into GDP-Gal. Careful biochemical analyses of the specific activity of recombinant Arabidopsis At5g28840 (494) and rice (497) protein demonstrate that the enzyme can convert GDP-Man to both GDP-Gul and GDP-Gal (496). A crystal structure of the enzyme was recently obtained (497).

5.6.19 CMP- β -KDO (CMP-KDO)

The eight-carbon acid sugar KDO, 3-deoxy-D-manno-2-octulosonic acid, is a primary sugar constituent in various types of cell surface extracellular polysaccharides and liposaccharides of Gram-negative bacteria. In plants, KDO is found only in RG-II. Synthesis of CMP-KDO in plants requires the activities of three enzymes.

- 1 KdsA, KDO-8-P synthase, catalyzes a condensation of PEP, phosphoenolpyruvate, and phosphorylated monosaccharide, D-arabinose 5-phosphate (A5P) in the presence of metal. Functional genes encoding KDO-8-P synthase activity were isolated from various plant species (498–500) and the encoded proteins share ~50% amino acid sequence identity with the bacterial proteins. In *Arabidopsis*, two gene isoforms (AtkdsA1, At1g79500; and AtkdsA2, At5g09730) were identified. The encoded isoforms share 93% amino acid sequence identity to each other. Interestingly, AtkdsA1 is predominantly expressed in shoots, while AtkdsA2 transcript accumulates to a higher level in roots. The activity of the recombinant plant kdsA toward other phosphorylated-sugars, such as D-erythrose-4-phosphate (E4P) was not tested. Based on bioinformatics, the plant KdsA are predicted to reside in the cytosol.
- 2 KDO-8-P phosphatase activity removes the phosphate to form KDO. The nature and specificity of this phosphatase is unknown.
- 3 KdsB, CMP-KDO synthase (CMP-KDOs), catalyzes the transfer of the cytidylyl group (CMP) from CTP to KDO in the production of the unusual nucleotide-sugar, CMP- β -KDO. The resulting activated sugar has a half-life of about 30 minutes in solution. The maize gene homolog was functionally identified (501), and the homologous *Arabidopsis* protein sequence is encoded by At1g53000. *In vitro*, the recombinant maize CMP-KDO is capable of using both CTP and UTP as nucleotide (Bar-Peled, unpublished). Interestingly, the plant proteins that share 40–50% amino acid sequence identity to the bacterial KdsA proteins have a 50 amino acid N-terminal extension. Bioinformatic analysis suggests that plant CMP-KDO is a Type Ib transmembrane protein with the catalytic domain predicted to face the cytosol. However, the subcellular location of the plant protein is uncertain. Since the function of RGII:2,3KDOT activity was not reported, it remains unclear if CMP or UMP-Kdo are the sugar donors.

5.6.20 Other enzymes involved in NDP-sugar metabolism

Feingold (403) summarized the NDP-sugars identified in plants. Those that were not described above are ADP-L-Ara, GDP-L-Ara; ADP-ribose; GDP-Xyl, ADP-Gal, GDP-D-Gal; ADP-D-Man; UDP-Fructose, ADP-D-Fructose; UDP-D-digitoxose (2,6-dideoxy-D-ribohexose); TDP-GalA; UDP-2-deoxy-2-acetamido-D-Glc, UDP-2-deoxy-2-acetamido-D-Gal; UDP-cellobiose. In addition to these NDP-sugars, pectin consists of other sugar residues: aceric acid and DHA (deoxyxyxoheptulopyranosilaric acid). The formation of these NDP-sugars is not well studied.

It is not clear if modifications of sugar residues on pectin (such as methylation or acetylation) occur after transferring the sugar from the respective NDP-sugar. In chloroplasts for example, sulfolipid biosynthesis requires the activated sugar UDP-Glc-6-sulfonate, UDP 5'-diphospho-sulfoquinovose. In this case, the sugar-linked to NDP is modified with a sulfate group prior to the transfer of the sulfoquinovose. Whether such NDP-sugar modifications occur with NDP-sugars required for wall synthesis is unknown. While more is known about synthesis of activated sugars less is known about "catabolism" of NDP-sugars. Recently, work on ascorbic acid metabolism in plants revealed two mutants, *vtc2* and *vtc4*, involved in the degradation of GDP-L-Gal. First, a GDP-L-Gal phosphorylase (*vtc2*, At4g26850) converts GDP-L-Gal and Pi into L-Gal-1-P and GDP (502) and subsequently a dephosphorylase

activity on L-Gal-1-P yields L-Gal (458) (vtc4, At3g02870). Interestingly, the phosphorylase converts GDP-D-Glc to D-Glc-1-P and GDP as well.

5.6.21 *Future questions and directions*

Currently, little is known about how the synthesis of nucleotide-sugars is controlled in time or space, and how it relates to the glycosyltransferases that actually make the diverse glycan polymers. What is the limiting factor in wall synthesis? Is it supply of NDP-sugars (as is the case for starch) or glycosyltransferases?

We will divide this section into three topics: sugar flux, role of isoforms, topology and protein complexes.

5.6.21.1 *Sugar flux*

Although a considerable proportion of cellular sugar ends up in wall polysaccharides, some sugar-derivatives are required for glycoprotein, glycolipid, and glycoside synthesis. In addition, significant amounts of sugars are stored either as large glycans such as starch, small-sized glycans (e.g., raffinose, fructan), or as the disaccharide sucrose. We would like to point out two issues related to flux: 1) growth potential of a cell; 2) whether some wall components compensate for the lack, or reduced amount, of other glycans.

- 1 New meristematic cells need to expand and grow to their prospective developmental tissue (e.g., leaf cells). What determines the growth potential and the cell's final size is unclear. Logically, with limited wall precursors the potential for growth is restricted since wall polymers are not made. The underlying mechanism that controls this complex developmental process is still unknown and poses a fascinating scientific quest. For example, do transcription factors regulate coordinately the expression of "tissue-fate genes" as well as NDP-sugar biosynthetic genes and genes involved in the supply of carbon? If carbon flux is not limited and all NDP-sugar biosynthetic genes are highly expressed – would the cell be larger? Do young, old, or stressed cells sense sugar availability or sugar status for growth and/or for storage in different ways? What are the ultimate determinates for growth; sugars or sugar-phosphates? Several sugar-sensing (signaling) proteins (and corresponding genes) have been isolated. It is assumed that sugar sensing (i.e., the interaction between a sugar molecule and a sensor protein) mediates a signal which initiates signal transduction cascades that result in cellular responses such as altered gene expression and enzymatic activities. Sugars as signaling molecules affect the plants at all stages of growth starting from seed germination to seed development. Sugars, like hormones, can act as primary messengers and regulate signals that control the expression of various genes involved in sugar-phosphates and wall metabolism. But do NDP-sugars function, in part, as signal molecules? In human cells, UDP-GlcNAc serves as a glucose sensor and moves between the cytosol and the nucleus. A cytosolic and nuclear-localized soluble enzyme, known as OGT, catalyzes the O-linked transfer of GlcNAc from UDP-GlcNAc directly to Ser/Thr of target proteins (503). For example, the O-GlcNAcylation, of the transcription factor Sp1 promotes nuclear localization of Sp1 and its ability to transactivate calmodulin (CaM) gene transcription (504). Whether plant cells consist of analogous signals to suppress or activate wall-biosynthetic genes by monitoring levels of NDP-sugars is unknown.

A major task for future research will be to investigate the relationships between isoforms that produce the same nucleotide-sugar, GTs, and sugar-sensing genes. Once the function of wall-related genes becomes known, bioinformatics will be useful in identifying a common set of genes that are coordinately expressed or suppressed to form a specific glycan.

5.6.21.2 *Role of isoforms*

Of course, the diversity of wall glycans could be more complex than suggested above. It is possible that certain cells will tolerate severe alterations in wall composition or will tolerate the complete lack of one or two types of wall glycans. The cell affected could either “reinforce” its wall with another glycan structure or not. If wall composition is flexible, then why do so many isoforms for the synthesis of the same NDP-sugar exist? One can argue that the existence of multiple NDP-sugar isoforms provides wall flexibility. A hypothetical example follows, reductions in the amounts of wall xylan, whose precursor UDP-xylose is produced by UXS Type A, can be overcome by the overproduction of xyloglucan, whose precursor UDP-Xyl is produced by UXS Type C. Indeed, genetic manipulation of a specific NDP-sugar (UGE) biosynthetic gene altered specific glycan synthesis (404).

One major consequence of the genome-sequencing project is the finding that what differentiates plants from other organisms (human, animal, yeast, or prokaryotes) is the fact that each NDP-sugar is synthesized by multiple isoforms. The current knowledge also raises many questions which were not apparent in the golden days of biochemistry. For example, why have plants evolved to generate so many isoforms for the synthesis of the same NDP-sugar? What is the evolutionary advantage for the synthesis of the same sugar attached to different nucleosides, for example, GDP-, CDP-, ADP-, UTP-, and TDP-glucose? Some species, for example bacteria, have a preference for TDP-, CDP-, and GDP-sugars, while vertebrates have preferences for UDP- and GDP-sugars. Bacteria, for example, produce TDP-Rha, whereas plants produce UDP-Rha. Such an activated sugar is not found in humans. If we keep in mind that all NDP-sugar biosynthetic genes likely evolved from one or several ancestral cells, and that the sequence similarity of all NDP-sugars is so high across species, one can form a testable theory for the rise of synonymous NDP-sugars. We will discuss possible roles of NDP-sugars using an analogy for the evolutionary pressure that presumably lead to the formation of 64 CODONS for the synthesis of 20 amino acids. It is now clear that codon usage is not random and among synonymous codons, some codons are used preferentially. It is even more fascinating that some species, for example *Drosophila*, have their own particular codon biases, and their usage differs significantly from those preferred by *E. coli*, yeast, or plants. Have NDP-sugars and diverse glycan structures evolved in the same manner? Why, for example, does one species, need GDP-Glc, ADP-Glc, CDP-Glc, and UDP-Glc. After all, it is the same Glc residue that ends up in the glycan (not the nucleoside). Would bacteria infecting humans or plants have an advantage in producing glycans using NDP-sugars not used by the host? After all, NDP-sugars are not imported to, or exported from, the infected cell.

One other possibility for an evolutionary pressure to produce multiple nucleotides bound to Glc, for example, can be explained by the Bar-Peled-Mohnen “Selfish Glycans Theory.” The diversity of glycans that each cell produces could not have successfully evolved if different GTs had to compete for the same NDP-sugar. If, for example, the biosynthetic apparatus responsible for the massive synthesis of starch, cellulose, or xyloglucans would all compete for UDP-Glc, the cell would either not grow or not be able to store carbohydrates. The “Selfish Glycan Theory” suggests that glycans are so essential for a living organism (just

like proteins) that each glycan requires its own supply of precursors. Extending this idea can be viewed by the specificity of GlycTs to their donors. It is well established that most, if not all, GlycTs are inhibited by the nucleotide-diphosphate. The data to date suggest that GlycT first binds the nucleotide then the sugar. Collectively, it would make “sense” that specific GlycTs avoid inhibition (by NDP) and compete for the sugar by selective pressure to adopt UDP-Glc rather than ADP-Glc, for example.

To make sure that glycans will succeed, eukaryotic cells also evolved to form these essential glycans in specific subcompartments. Starch, for example, is produced in plastids, whereas xyloglucan is made in the Golgi and cellulose at the plasma membrane. To accommodate synthesis of these glycans, multiple isoforms that synthesize either the same NDP-Glc (UDP-Glc) or different NDP-Glc (ADP-Glc) evolved for each glycan. Of course, it will also make sense that plants would produce the minimum amount of these NDP-sugars inside the cell to avoid metabolic waste and to efficiently incorporate these precursors in the appropriate subcellular sites where each glycan is made.

5.6.21.3 *Cellular location and enzyme topology*

DNA is a water-soluble molecule, aided by tens of proteins to helping in the packaging and re-packing during RNA and DNA synthesis. Proteins are also soluble and if they are too hydrophobic, either a portion will be embedded in membranes or hydrophobic domains will assemble together with other hydrophobic proteins to maintain their solubility. On the other hand, plant polysaccharides provide us with a challenge in terms of understanding how such massive amounts of insoluble or gel-like structures are made inside the Golgi apparatus without hindering other cellular processes. Bacteria, for example, form insoluble lipo-glycans in stages: UDP-sugars made in the cytosol contribute to synthesis of short side chain glycans attached to the inner membrane. These short glycans are made facing the cytosol. The side chains are then flipped and transferred through the outer membrane to the outside of the cell, where specific enzymes cut and assemble the side chains to form complete glycan structures. This process, by analogy, is very similar to the synthesis of core N-linked glycoproteins. But how are pectins and hemicelluloses made in plants? Are Golgi-synthesized polysaccharides made completely in the Golgi, or are they made in smaller fragments that are assembled together at the wall, as is the case in lipo-glycan synthesis? If the entire pectic polymers are made in the Golgi, then are they sequestered or packed temporarily with or by proteins to help with the challenge posed by their physical properties (i.e., solubility, size, etc.)? Further questions will arise. Is HG made inside the lumen (thus making a jelly-like lumen) or perhaps made in specific Golgi stacks, designated only for the wall (not glycoprotein biosynthesis)? The latter scenario could be attractive in light of the fact that unlike humans and fungi, one plant cell consists of hundreds of Golgi stacks. It is possible that an entire “designated Golgi” is moving with its packed glycan(s) to the wall to “unload” the insoluble matrix.

While many of the putative GTs are Type II membrane proteins (i.e., predicted to have their catalytic domain facing the lumen), some GTs, such as mannan synthase, have multi domains that span the membranes. Where is the catalytic domain of such Golgi-enzymes facing? Furthermore, if a glycan is fully made in the Golgi, is it made in one subcompartment of the Golgi, for example, cis-Golgi, or is it initiated in the cis- and further modified in the medial-Golgi (like glycoproteins)? Certainly, understanding where each wall biosynthetic enzyme functions and where its catalytic domain faces is essential to address the above questions.

In addition to location and topology, it is puzzling how polysaccharides are made in the small Golgi cisternae (estimated to be smaller than 20×200 nm). The same cisternae are temporarily packed not only with wall-glycans but also with numerous Golgi-resident proteins and large numbers of secretory proteins. If the average size of the catalytic domain of GTs and membrane-bound NDP-sugar biosynthetic enzymes is 30–40 kDa, they can “touch” each other if inserted opposite to each other. Given the low quantity of these metabolic enzymes and the potential solubility problems of some glycans, one can wonder if synthesis of a specific Golgi-glycan is done in a complex of enzymes, as is the case in the synthesis of cellulose.

5.7 Perspectives

Traditional protein biochemistry and classical purification was instrumental in identifying key wall biosynthetic enzymes, for example, XG:1-2,fucosyltransferase (116, 117) and GM:1-6,galactosyltransferase (88). As most of the glycosyltransferases involved in polysaccharide synthesis are membrane bound and are of low abundance, i.e., estimated 2–6 molecules per cell, it is clear that other methods should be sought to identify wall biosynthetic proteins. Classical genetics has already contributed immensely to the identification of large numbers of genes involved in glycan synthesis such as cellulose synthase and their large gene family (507, 508) and xyloglucan synthesis (120). The genome sequence facilitated the identification of mannan synthase (85) and cellulose-like proteins (154). A combined approach of partial protein purification and proteomics was useful in the identification of a large family associated with the pectin biosynthetic HG:galacturonosyltransferases (137). The classification of GTs in the CAZY database was instrumental in identifying GT gene candidates targeted for a reverse genetic approach using the SALK T-DNA or other mutant library collections. Lastly, microarray analyses of tissues or cell types at different developmental stages were useful in identifying secondary wall synthesis candidates (139, 509). Similar biochemical, genetic, and genomic approaches were successful in identifying biosynthetic genes involved in nucleotide-sugar synthesis and NDP-sugar transporters. The last decade has been a very fruitful and exciting time for the community of wall researchers (finally a crack in the wall).

Can bioinformatics assist in the prediction of GTs? Would the knowledge of UDP-binding sites in several plant GTs be useful to distinguish between GDP-, or ADP-transferases? Would the knowledge of the sugar moiety-binding site be similarly helpful to identify, for example, putative GalTs? Is the binding pocket for UDP-Rha in a flavonoid: RhaT the same as for pectin:RhaT? Clearly, a critical mass of biochemical knowledge is required to start to predict gene function by computer. We are not there yet. . . .

Acknowledgments

A review of this magnitude is not possible without the input, efforts, and patience of countless individuals. The authors express sincere gratitude to the multiple students, researchers, colleagues, and staff who contributed to this chapter through engaging discussions and editorial effort. Special thanks go out to our families who graciously endured our too frequent absences. We also thank the Department of Energy, the National Science Foundation and

the NRI, CSREES, USDA who provide the invaluable funding that supports plant cell wall research.

References

1. Richmond, T. & Somerville, C.R. (2001) Integrative approaches to determining Csl function. *Plant Molecular Biology*, **47**, 131–143.
2. Albersheim, P., Darvill, A.G., O'Neill, M.A., Schols, H.A. & Voragen, A.G.J. (1996) An hypothesis: The same six polysaccharides are components of the primary cell walls of all higher plants. In: *Pectins and Pectinases*, Vol. 14 (eds. J. Visser & A.G.J. Voragen), pp. 47–53. Elsevier Sciences B.V., Amsterdam.
3. Somerville, C. (2006) Cellulose synthesis in higher plants. *Annual Review of Cell and Developmental Biology*, **22**, 53–78.
4. Turner, S.R. & Somerville, C.R. (1997) Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. *Plant Cell*, **9**, 689–701.
5. Jarvis, M.C. (2000) Interconversion of the Ialpha and Ibeta crystalline forms of cellulose by bending. *Carbohydrate Research*, **325**, 150–154.
6. Brett, C.T. (2000) Cellulose microfibrils in plants: Biosynthesis, deposition, and integration into the cell wall. In: *International Review of Cytology*, Vol. 199 (ed. K.W. Jeon), pp. 161–199. Academic Press, London UK.
7. Brown, R.M., Jr. (2004) Cellulose structure and biosynthesis: What is in store for the 21st century? *Journal of Polymer Science A*, **42**, 487–495.
8. Lane, D.R., Wiedemeier, A., Peng, L.C., Höfte, H., Vernhettes, S., Desprez, T., Hocart, C.H., Birch, R.J., Baskin, T.I., Burn, J.E., Arioli, T., Betzner, A.S. & Williamson, R.E. (2001) Temperature-sensitive alleles of RSW2 link the KORRIGAN endo-1;4-beta-glucanase to cellulose synthesis and cytokinesis in Arabidopsis. *Plant Physiology*, **126**, 278–288.
9. Römling, U. (2002) Molecular biology of cellulose production in bacteria. *Research Microbiology*, **153**, 205–212.
10. Ha, M.A., Apperley, D.C., Evans, B.W., Huxham, I.M., Jardine, W.G., Vietor, R.J., Reis, D., Vian, B. & Jarvis, M.C. (1998) Fine structure in cellulose microfibrils: NMR evidence from onion and quince. *Plant Journal*, **16**, 183–190.
11. Herth, W. (1983) Arrays of plasma-membrane “rosettes” involved in cellulose microfibril formation of *Spirogyra*. *Planta*, **159**, 347–356.
12. Kimura, S., Laosinchai, W., Itoh, T., Cui, X.J., Linder, C.R. & Brown, R.M. (1999) Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. *Plant Cell*, **11**, 2075–2085.
13. Holland, N., Holland, D., Helentjaris, T., Dhugga, K.S., Xoconostle-Cazares, B. & Delmer, D.P. (2000) A comparative analysis of the plant cellulose synthase (CesA) gene family. *Plant Physiology*, **123**, 1313–1323.
14. Richmond, T. (2000) Higher plant cellulose synthases. *Genome Biology*, **1**(4), 3001.1–3001.6.
15. Saurin, A.J., Borden, K.L., Boddy, M.N. & Freemont, P.S. (1996) Does this have a familiar RING? *Trends in Biochemical Sciences*, **21**, 208–214.
16. Saxena, I.M. & Brown, R.M., Jr. (1997) Identification of cellulose synthase(s) in higher plants: Sequence analysis of processive beta-glycosyltransferases with the common motif “D; D; D35Q(R; Q)XRW”. *Cellulose*, **4**, 33–49.
17. Marks, D.L., Dominguez, M., Wu, K.J. & Pagano, R.E. (2001) Identification of active site residues in glucosylceramide synthase: A nucleotide-binding/catalytic motif conserved with processive beta-glycosyltransferases. *Journal of Biological Chemistry*, **276**, 26492–26498.

18. Nagahashi, S., Sudoh, M., Ono, N., Sawada, R., Yamaguchi, E., Uchida, Y., Mio, T., Takagi, M., Arisawa, M. & Yamada-Okabe, H. (1995) Characterization of chitin synthase 2 of *Saccharomyces cerevisiae*: Implication of two highly conserved domains as possible catalytic sites. *Journal of Biological Chemistry*, **270**, 13961–13967.
19. Saxena, I.M., Brown, R.M., Jr. & Dandekar, T. (2001) Structure–function characterization of cellulose synthase: Relationship to other glycosyltransferases. *Phytochemistry*, **57**, 1135–1148.
20. Taylor, N.G., Laurie, S. & Turner, S.R. (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. *Plant Cell*, **12**, 2529–2539.
21. Taylor, N.G., Howells, R.M., Huttly, A.K., Vickers, K. & Turner, S.R. (2003) Interactions among three distinct CesaA proteins essential for cellulose synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 1450–1455.
22. Kudlicka, K., Brown, R.M. Jr., Li, L., Lee, J., Shin, H. & Kuga, S. (1995) β -glucan synthesis in the cotton fiber. IV In vitro assembly of the cellulose I allomorph. *Plant Physiology*, **107**, 111–123.
23. Kudlicka, K. & Brown, R.M., Jr. (1997) Cellulose and callose biosynthesis in higher plants. I. Solubilization and separation of (1,3)- and (1,4)-beta-glucan synthase activities from mung bean. *Plant Physiology*, **115**, 643–656.
24. Lai-Kee-Him, J., Chanzy, H., Muller, M., Putaux, J.L., Imai, T. & Bulone, V. (2002) In vitro versus in vivo cellulose microfibrils from plant primary wall synthases: Structural differences. *Journal of Biological Chemistry*, **277**, 36931–36939.
25. Ohlsson, A.B., Djerbi, S., Winzell, A., Bessueille, L., Ståldal, V., Li, X., Blomqvist, K., Bulone, V., Teeri, T.T. & Berglund, T. (2006) Cell suspension cultures of *Populus tremula* X *P. tremuloides* exhibit a high level of cellulose synthase gene expression that coincides with increased in vitro cellulose synthase activity. *Protoplasma*, **228**, 221–229.
26. Koyama, M., Helbert, W., Imai, T., Sugiyama, J. & Henrissat, B. (1997) Parallel-up structure evidences the molecular directionality during biosynthesis of bacterial cellulose. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9091–9095.
27. Han, N.S. & Robyt, J.F. (1998) Mechanism of *Acetobacter xylinum* cellulose biosynthesis: Direction of chain elongation and the role of lipid pyrophosphate intermediates in the cell membrane. *Carbohydrate Research*, **313**, 125–133.
28. Sugiyama, J., Boisset, C., Hashimoto, M. & Watanabe, T. (1999) Molecular directionality of beta-chitin biosynthesis. *Journal of Molecular Biology*, **286**, 247–255.
29. Unligil, U.M. & Rini, J.M. (2000) Glycosyltransferase structure and mechanism. *Current Opinion in Structural Biology*, **10**, 510–517.
30. Carpita, N. & Vergara, C. (1998) A recipe for cellulose. *Science*, **279**, 672–673.
31. Yeager, A.R. & Finney, N.S. (2004) The first direct evaluation of the two-active site mechanism for chitin synthase. *Journal of Organic Chemistry*, **69**, 613–618.
32. Chang, R., Yeager, A.R. & Finney, N.S. (2003) Probing the mechanism of a fungal glycosyltransferase essential for cell wall biosynthesis. UDP-Chitobiose is not a substrate for chitin synthase. *Organic Biomolecular Chemistry*, **1**, 39–41.
33. Peng, L., Xiang, F., Roberts, E., Kawagoe, Y., Greve, L.C., Kreuz, K. & Delmer, D.P. (2001) The experimental herbicide CGA 325'615 inhibits synthesis of crystalline cellulose and causes accumulation of non-crystalline beta-1;4-glucan associated with CesaA protein. *Plant Physiology*, **126**, 981–992.
34. Franco, O.L. & Rigden, D.J. (2003) Fold recognition analysis of glycosyltransferase families: Further members of structural superfamilies. *Glycobiology*, **13**, 707–712.
35. Peng, L., Kawagoe, Y., Hogan, P. & Delmer, D. (2002) Sitosterol-beta-glucoside as primer for cellulose synthesis in plants. *Science*, **295**, 147–150.

36. Ihara, Y., Takeda, T., Sakai, F. & Hayashi, T. (2002) Transferase activity of GhCesA2 (putative cotton cellulose 4-, β -glucosyltransferase) expressed in *Pichia pastoris*. *Journal of Wood Science*, **48**, 425–428.
37. Shaw, S.L., Kamyar, R. & Ehrhardt, D.W. (2003) Sustained microtubule treadmilling in *Arabidopsis* cortical arrays. *Science*, **300**, 1715–1718.
38. Ledbetter, M. & Porter, K. (1963) A microtubule in plant cell fine structure. *Journal of Cell Biology*, **19**, 239–250.
39. Baskin, T.I. (2001) On the alignment of cellulose microfibrils by cortical microtubules: A review and a model. *Protoplasma*, **215**, 150–171.
40. Giddings, T.H., Jr. & Staehelin, L.A. (1991) Microtubule-mediated control of microfilament deposition: A re-examination of the hypothesis. In: *The Cytoskeletal Basis of Plant Growth and Form* (ed. C.W. Lloyd), pp. 85–99. Academic Press, New York.
41. Alberts, B., Johnson, A., Lewis, J.S., Raff, M., Roberts, K. & Walter, P. (2002) *Molecular Biology of the Cell*. Garland, New York.
42. Wasteneys, G.O. (2004) Progress in understanding the role of microtubules in plant cells. *Current Opinion in Plant Biology*, **7**, 651–660.
43. Sugimoto, K., Himmelsbach, R., Williamson, R.E. & Wasteneys, G.O. (2003) Mutation or drug-dependent microtubule disruption causes radial swelling without altering parallel cellulose microfibril deposition in *Arabidopsis* root cells. *Plant Cell*, **15**, 1414–1429.
44. Baskin, T.I., Beemster, G.T.S., Judy-March, J.E. & Marga, F. (2004) Disorganization of cortical microtubules stimulates tangential expansion and reduces the uniformity of cellulose microfibril alignment among cells in the root of *Arabidopsis*. *Plant Physiology*, **135**, 2279–2290.
45. Sugimoto, K., Williamson, R.E. & Wasteneys, G.O. (2000) New techniques enable comparative analysis of microtubule orientation; wall texture; and growth rate in intact roots of *Arabidopsis*. *Plant Physiology*, **124**, 1493–1506.
46. Himmelsbach, R., Williamson, R.E. & Wasteneys, G.O. (2003) Cellulose microfibril alignment recovers from DCB-induced disruption despite microtubule disorganization. *Plant Journal*, **36**, 565–575.
47. Paredez, A., Wright, A. & Ehrhardt, D.W. (2006) Microtubule cortical array organization and plant cell morphogenesis. *Current Opinion in Plant Biology*, **9**, 571–578.
48. Haigler, C.H. & Brown, R.M., Jr. (1986) Transport of rosettes from the Golgi apparatus to the plasma membrane in isolated mesophyll cells of *Zinnia-elegans* during differentiation to tracheary elements in suspension culture. *Protoplasma*, **134**, 111–120.
49. Hirai, N., Sonobe, S. & Hayashi, T. (1998) In situ synthesis of beta-glucan microfibrils on tobacco plasma membrane sheets. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 15102–15106.
50. Valdivia, R.H. & Schekman, R. (2003) The yeasts Rho1p and Pkc1p regulate the transport of chitin synthase III (Chs3p) from internal stores to the plasma membrane. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 10287–10292.
51. Rudolph, U. & Schnepf, E. (1988) Investigations of the turnover of the putative cellulose-synthesizing particle rosettes within the plasma membrane of *Funaria hygrometrica* protonema cells. I. Effects of monensin and cytochalasin B. *Protoplasma*, **143**, 63–73.
52. DeBolt, S., Gutierrez, R., Ehrhardt, D.W., Melo, C.V., Ross, L., Cutler, S.R., Somerville, C. & Bonetta, D. (2007) Morlin, an inhibitor of cortical microtubule dynamics and cellulose synthase movement. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5854–5859.
53. Emons, A.M.C. & Mulder, B.M. (1998) The making of the architecture of the plant cell wall: How cells exploit geometry. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 7215–7219.

54. Gardiner, J.C., Taylor, N.G. & Turner, S.R. (2003) Control of cellulose synthase complex localization in developing xylem. *Plant Cell*, **15**, 1740–1748.
55. Fisher, D.D. & Cyr, R.J. (1998) Extending the microtubule/microfibril paradigm – Cellulose synthesis is required for normal cortical microtubule alignment in elongating cells. *Plant Physiology*, **116**, 1043–1051.
56. Lazzaro, M.D., Donohue, J.M. & Soodavar, F.M. (2003) Disruption of cellulose synthesis by isoxaben causes tip swelling and disorganizes cortical microtubules in elongating conifer pollen tubes. *Protoplasma*, **220**, 201–207.
57. Römmling, U., Gomelsky, M. & Galperin, M.Y. (2005) C-di-GMP: The dawning of a novel bacterial signalling system. *Molecular Microbiology*, **57**, 629–639.
58. Amor, Y., Mayer, R., Benziman, M. & Delmer, D. (1991) Evidence for a cyclic diguanylic acid-dependent cellulose synthase in plants. *Plant Cell*, **3**, 989–995.
59. Nuhse, T.S., Stensballe, A., Jensen, O.N. & Peck, S.C. (2004) Phosphoproteomics of the arabidopsis plasma membrane and a new phosphorylation site database. *Plant Cell*, **16**, 2394–2405.
60. Taylor, N.G. (2007) Identification of cellulose synthase AtCesA7 (IRX3) in vivo phosphorylation sites – a potential role in regulating protein degradation. *Plant Molecular Biology*, **64**, 161–171.
61. Haigler, C.H., Ivanova-Datcheva, M., Hogan, P.S., Salnikov, V.V., Hwang, S., Martin, K. & Delmer, D.P. (2001) Carbon partitioning to cellulose synthesis. *Plant Molecular Biology*, **47**, 29–51.
62. Amor, Y., Haigler, C.H., Johnson, S., Wainscott, M. & Delmer, D.P. (1995) A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 9353–9357.
63. Bieniawska, Z., Barratt, D.H.P., Garlick, A.P., Thole, V., Kruger, N.J., Martin, C., Zrenner, R. & Smith, A.M. (2007) Analysis of the sucrose synthase gene family in Arabidopsis. *Plant Journal*, **49**, 810–828.
64. Ruan, Y.L., Llewellyn, D.J. & Furbank, R.T. (2003) Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. *Plant Cell*, **15**, 952–964.
65. Coleman, H.D., Ellis, D.D., Gilbert, M. & Mansfield, S.D. (2006) Up-regulation of sucrose synthase and UDP-glucose pyrophosphorylase impacts plant growth and metabolism. *Plant Biotechnology Journal*, **4**, 87–101.
66. Hamann, T., Osborne, E., Youngs, H.L., Misson, J., Nussaume, L. & Somerville, C. (2004) Global expression analysis of CESA and CSL genes in Arabidopsis. *Cellulose*, **11**, 279–286.
67. Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. & Gruissem, W. (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiology*, **136**, 2621–2632.
68. Beeckman, T., Przemeck, G.K.H., Stamatiou, G., Lau, R., Terryn, N., De Rycke, R., Inze, D. & Berleth, T. (2002) Genetic complexity of cellulose synthase A gene function in Arabidopsis embryogenesis. *Plant Physiology*, **130**, 1883–1893.
69. Burton, R.A., Shirley, N.J., King, B.J., Harvey, A.J. & Fincher, G.B. (2004) The CesA gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. *Plant Physiology*, **134**, 224–236.
70. Turner, S.R., Taylor, N. & Jones, L. (2001) Mutations of the secondary cell wall. *Plant Molecular Biology*, **47**, 209–219.
71. Appenzeller, L., Doblin, M., Barreiro, R., Wang, H., Niu, X., Kollipara, K., Carrigan, L., Tomes, D., Chapman, M. & Dhugga, K.S. (2004) Cellulose synthesis in maize: Isolation and expression analysis of the cellulose synthase (CesA) gene family. *Cellulose*, **11**, 287–299.
72. Dhugga, K.S. (2001) Building the wall: Genes and enzyme complexes for polysaccharide synthases. *Current Opinion in Plant Biology*, **4**, 488–493.
73. Brown, D.M., Zeef, L.A.H., Ellis, J., Goodacre, R. & Turner, S.R. (2005) Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell*, **17**, 2281–2295.

74. Persson, S., Wei, H., Milne, J., Page, G.P. & Somerville, C.R. (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8633–8638.
75. Hu, W.J., Harding, S.A., Lung, J., Popko, J.L., Ralph, J., Stokke, D.D., Tsai, C.J. & Chiang, V.L. (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nature Biotechnology*, **17**, 808–812.
76. Zenoni, S., Reale, L., Tornielli, G.B., Lanfalone, L., Porceddu, A., Ferrarini, A., Moretti, C., Zamboni, A., Speghini, A., Ferranti, F. & Pezzotti, M. (2004) Downregulation of the *Petunia hybrida* α -expansin gene *PhEXP1* reduces the amount of crystalline cellulose in cell walls and leads to phenotypic changes in petal limbs. *Plant Cell*, **16**, 295–308.
77. Marga, F., Grandbois, M., Cosgrove, D.J. & Baskin, T.I. (2005) Cell wall extension results in the coordinate separation of parallel microfibrils: Evidence from scanning electron microscopy and atomic force microscopy. *Plant Journal*, **43**, 181–190.
78. Elbein, A.D. (1969) Biosynthesis of a cell wall glucomannan in mung bean seedlings. *Journal of Biological Chemistry*, **244**, 1608–1616.
79. Dalessandro, G., Piro, G. & Northcote, D.H. (1986) Glucomannan synthase activity in differentiating cells of *Pinus sylvestris* L. *Planta*, **169**, 564–574.
80. Dunn, E.K., Shoue, D.A., Huang, X., Kline, R.E., MacKay, A.L., Carpita, N.C., Taylor, I.E. & Mandoli, D.F. (2007) Spectroscopic and biochemical analysis of regions of the cell wall of the unicellular “mannan weed”, *Acetabularia acetabulum*. *Plant and Cell Physiology*, **48**, 122–133.
81. Liepman, A.H., Nairn, C.J., Willats, W.G.T., Sorensen, I., Roberts, A.W. & Keegstra, K. (2007) Functional genomic analysis supports conservation of function among cellulose synthase-like A gene family members and suggests diverse roles of mannans in plants. *Plant Physiology*, **143**, 1881–1893.
82. Edwards, M., Bulpin, P.V., Dea, I.C.M. & Reid, J.S.G. (1989) Biosynthesis of legume-seed galactomannans in-vitro: cooperative interactions of a guanosine 5'-diphosphate mannose-linked 1-4-beta-D mannosyltransferase and a uridine 5'-diphosphate galactose-linked alpha-D galactosyltransferase in particulate enzyme preparations from developing endosperms of fenugreek *Trigonella foenum graecum* L. and guar *Cyamopsis tetragonoloba* L. Taub. *Planta*, **178**, 41–51.
83. Piro, G., Zuppa, A., Dalessandro, G. & Northcote, D.H. (1993) Glucomannan synthesis in pea epicotyls: The mannose and glucose transferases. *Planta*, **190**, 206–220.
84. Schroder, R., Wegrzyn, T.F., Bolitho, K.M. & Redgwell, R.J. (2004) Mannan transglycosylase: A novel enzyme activity in cell walls of higher plants. *Planta*, **219**, 590–600.
85. Dhugga, K.S., Barreiro, R., Whitten, B., Stecca, K., Hazebroek, J., Randhawa, G.S., Dolan, M., Kinney, A.J., Tomes, D., Nichols, S. & Anderson, P. (2004) Guar seed β -mannan synthase is a member of the cellulose synthase super gene family. *Science*, **303**, 363–366.
86. Liepman, A.H., Wilkerson, C.G. & Keegstra, K. (2005) Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 2221–2226.
87. Suzuki, S., Li, L., Sun, Y.H. & Chiang, V.L. (2006) The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. *Plant Physiology*, **142**, 1233–1245.
88. Edwards, M.E., Dickson, C.A., Chengappa, S., Sidebottom, C., Gidley, M.J. & Reid, J.S.G. (1999) Molecular characterisation of a membrane-bound galactosyltransferase of plant cell wall matrix polysaccharide biosynthesis. *Plant Journal*, **19**, 691–697.
89. Reid, J.S.G., Edwards, M., Gidley, M.J. & Clark, A.H. (1995) Enzyme specificity in galactomannan biosynthesis. *Planta*, **195**, 489–495.
90. Handford, M.G., Baldwin, T.C., Goubet, F., Prime, T.A., Miles, J., Yu, X. & Dupree, P. (2003) Localisation and characterisation of cell wall mannan polysaccharides in *Arabidopsis thaliana*. *Planta*, **218**, 27–36.

91. Goubet, F., Misrahi, A., Park, S.K., Zhang, Z., Twell, D. & Dupree, P. (2003) AtCSLA7, a cellulose synthase-like putative glycosyltransferase, is important for pollen tube growth and embryogenesis in Arabidopsis. *Plant Physiology*, **131**, 547–557.
92. Zhu, Y., Nam, J., Carpita, N.C., Matthysse, A.G. & Gelvin, S.B. (2003) Agrobacterium-mediated root transformation is inhibited by mutation of an arabidopsis cellulose synthase-like gene. *Journal of Plant Physiology*, **133**, 1000–1010.
93. Baldwin, T.C., Handford, M.G., Yuseff, M.-I., Orellana, A. & Dupree, P. (2001) Identification and characterization of GONST1, a Golgi-localized GDP-mannose transporter in Arabidopsis. *Plant Cell*, **13**, 2283–2295.
94. Hayashi, T. (1989) Xyloglucans in the primary cell wall. *Annual Review of Plant Physiology*, **40**, 139–168.
95. Hayashi, T. & MacLachlan, G. (1984) Pea xyloglucan and cellulose I: Macromolecular organization. *Plant Physiology*, **75**, 596–604.
96. Taylor, I.E.P. & Atkins, E.D.T. (1985) X-ray-diffraction studies on the xyloglucan from tamarind (*Tamarindus-indica*) seed. *FEBS Letters*, **181**, 300–302.
97. Fry, S.C., York, W.S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J.P., Kato, Y., Lorences, E.P., MacLachlan, G.A., McNeil, M., Mort, A.J., Reid, J.S.G., Seitz, H.U., Selvendran, R.R., Voragen, A.G.J. & White, A.R. (1993) An unambiguous nomenclature for xyloglucan-derived oligosaccharides. *Physiologica Plantarum*, **89**, 1–3.
98. York, W.S., Kolli, V.S.K., Orlando, R., Albersheim, P. & Darvill, A.G. (1996) The structures of arabinoxylglucans produced by solanaceous plants. *Carbohydrate Research*, **285**, 99–128.
99. Lerouxel, O., Cavalier, D.M., Liepman, A.H. & Keegstra, K. (2006) Biosynthesis of plant cell wall polysaccharides – a complex process. *Current Opinion in Plant Biology*, **9**, 621–630.
100. Pauly, M., Albersheim, P., Darvill, A. & York, W.S. (1999) Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. *Plant Journal*, **20**, 629–639.
101. Otegui, M. & Staehelin, L.A. (2000) Syncytial-type cell plates: A novel kind of cell plate involved in endosperm cellularization of Arabidopsis. *Plant Cell*, **12**, 933–947.
102. Hoffman, M., Jia, Z., Peña, M.J., Cash, M., Harper, A., Blackburn, A., Darvill, A. & York, W.S. (2005) Structural analysis of xyloglucans in the primary cell walls of plants in the subclass *Asteridae*. *Carbohydrate Research*, **340**, 1826–1840.
103. York, W.S., Oates, J.E., Van Halbeek, H., Darvill, A.G., Albersheim, P., Tiller, P.R. & Dell, A. (1988) Location of the O-acetyl substituents on a nonasaccharide repeating unit of sycamore extracellular xyloglucan. *Carbohydrate Research*, **173**, 113–132.
104. Jia, Z., Cash, M., Darvill, A.G. & York, W.S. (2005) NMR characterization of endogenously O-acetylated oligosaccharides isolated from tomato (*Lycopersicon esculentum*) xyloglucan. *Carbohydrate Research*, **340**, 1818–1825.
105. Perrin, R.M., Jia, Z., Wagner, T.A., O'Neill, M.A., Sarria, R., York, W.S., Raikhel, N.V. & Keegstra, K. (2003) Analysis of xyloglucan fucosylation in arabidopsis. *Plant Physiology*, **132**, 768–778.
106. Moore, P.J. & Staehelin, L.A. (1988) Immunogold localization of the cell-wall-matrix polysaccharides rhamnogalacturonan I and xyloglucan during cell expansion and cytokinesis in *Trifolium pratense* L. Implication for secretory pathways. *Planta*, **174**, 433–445.
107. Thompson, J.E. & Fry, S.C. (2000) Evidence for covalent linkage between xyloglucan and acidic pectins in suspension-cultured rose cells. *Planta*, **211**, 275–286.
108. Brett, C.T., Baydoun, E.A.H. & Abdel-Massih, R.M. (2005) Pectin-xyloglucan linkages in type I primary cell walls of plants. *Plant Biosystems*, **139**, 54–59.
109. Kerr, E.M. & Fry, S.C. (2004) Extracellular cross-linking of xylan and xyloglucan in maize cell-suspension cultures: The role of oxidative phenolic coupling. *Planta*, **219**, 73–83.
110. Hayashi, T., Marsden, M.P.F. & Delmer, D.P. (1987) Pea xyloglucan and cellulose: V. Xyloglucan-cellulose interactions in vitro and in vivo. *Plant Physiology*, **83**, 384–389.

111. Levy, S., MacLachlan, G. & Staehelin, L.A. (1997) Xyloglucan sidechains modulate binding to cellulose during *in vitro* binding assays as predicted by conformational dynamics simulations. *Plant Journal*, **11**, 373–386.
112. McCann, M.C., Wells, B. & Roberts, K. (1990) Direct visualization of cross-links in the primary plant-cell wall. *Journal of Cell Science*, **96**, 323–334.
113. McCann, M.C., Wells, B. & Roberts, K. (1992) Complexity in the spatial localization and length distribution of plant cell-wall matrix polysaccharides. *Journal of Microscopy*, **166**, 123–136.
114. Fujino, T., Sone, Y., Mitsuishi, Y. & Itoh, T. (2000) Characterization of cross-links between cellulose microfibrils, and their occurrence during elongation growth in pea epicotyl. *Plant and Cell Physiology*, **41**, 486–494.
115. Vissenberg, K., Fry, S.C., Pauly, M., Höfte, H. & Verbelen, J.P. (2005) XTH acts at the microfibril-matrix interface during cell elongation. *Journal of Experimental Botany*, **56**, 673–683.
116. Perrin, R.M., DeRocher, A.E., Bar-Peled, M., Zeng, W., Norambuena, L., Orellana, A., Raikhel, N.V. & Keegstra, K. (1999) Xyloglucan fucosyltransferase, an enzyme involved in plant cell wall biosynthesis. *Science*, **284**, 1976–1979.
117. Faik, A., Bar-Peled, M., DeRocher, A.E., Zeng, W., Perrin, R.M., Wilkerson, C., Raikhel, N.V. & Keegstra, K. (2000) Biochemical characterization and molecular cloning of an α -1,2-fucosyltransferase that catalyzes the last step of cell wall xyloglucan biosynthesis in pea. *Journal of Biological Chemistry*, **275**, 15082–15089.
118. Perrin, R., Wilkerson, C. & Keegstra, K. (2001) Golgi enzymes that synthesize plant cell wall polysaccharides: Finding and evaluating candidates in the genomic era. *Plant Molecular Biology*, **47**, 115–130.
119. Vanzin, G.F., Madson, M., Carpita, N.C., Raikhel, N.V., Keegstra, K. & Reiter, W.D. (2002) The *mur2* mutant of *Arabidopsis thaliana* lacks fucosylated xyloglucan because of a lesion in fucosyltransferase AtFUT1. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 3340–3345.
120. Madson, M., Dunand, C., Li, X., Verma, R., Vanzin, G.A., Calplan, J., Shoue, D.C., Carpita, N.C. & Reiter, W.D. (2003) The *MUR3* gene of *Arabidopsis* encodes a xyloglucan galactosyltransferase that is evolutionarily related to animal exostosins. *Plant Cell*, **15**, 1662–1670.
121. Li, X., Cordero, I., Caplan, J., Molhoj, M. & Reiter, W.D. (2004) Molecular analysis of 10 coding regions from *Arabidopsis* that are homologous to the *MUR3* xyloglucan galactosyltransferase. *Plant Physiology*, **134**, 940–950.
122. Faik, A., Price, N.J., Raikhel, N.V. & Keegstra, K. (2002) An *Arabidopsis* gene encoding an α -xylosyltransferase involved in xyloglucan biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 7797–7802.
123. White, A.R., Xin, Y. & Pezeshk, V. (1993) Xyloglucan glucosyltransferase in Golgi membranes from *Pisum sativum* (pea). *Biochemical Journal*, **294**, 231–238.
124. Cocuron, J.C., Lerouxel, O., Drakakaki, G., Alonso, A.P., Liepman, A.H., Keegstra, K., Raikhel, N. & Wilkerson, C.G. (2007) A gene from the cellulose synthase-like C family encodes a β -1,4 glucan synthase. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 8550–8555.
125. Reiter, W.D., Chapple, C. & Somerville, C.R. (1997) Mutants of *Arabidopsis thaliana* with altered cell wall polysaccharide composition. *Plant Journal*, **12**, 335–345.
126. Reiter, W.D. (2002) Biosynthesis and properties of the plant cell wall. *Current Opinion in Plant Biology*, **5**, 536–542.
127. Ryden, P., Sugimoto-Shirasu, K., Smith, A.C., Findlay, K., Reiter, W.D. & McCann, M.C. (2003) Tensile properties of *Arabidopsis* cell walls depend on both a xyloglucan cross-linked microfibrillar network and rhamnogalacturonan II-borate complexes. *Plant Physiology*, **132**, 1033–1040.

128. Ebringerová, A. & Heinze, T. (2000) Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties. *Macromolecular Rapid Communications*, **21**, 542–556.
129. Peña, M., Zhong, R., Zhou, G.-K., Richardson, E., O'Neill, M.A., Darvill, A.G., York, W.S. & Ye, Z.-H. (2007) Alterations in the abundance of the glycosyl sequence at the reducing end of glucuronoxylan and changes in glucuronoxylan chain length in the *Arabidopsis irregular xylem8* and 9mutants. *Plant Cell*, **19**, 549–563.
130. Zhong, R., Peña, M.J., Zhou, G.K., Nairn, C.J., Wood-Jones, A., Richardson, E.A., Morrison III, W.H., Darvill, A.G., York, W.S. & Ye, Z.F.H. (2005) Arabidopsis Fiber 8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. *Plant Cell*, **17**, 3390–3408.
131. Bauer, S., Vasu, P., Persson, S., Mort, A.J. & Somerville, C.R. (2006) Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 11417–11422.
132. Persson, S., Caffall, K.H., Freshour, G., Hilley, M.T., Bauer, S., Poindexter, P., Hahn, M.G., Mohnen, D. & Somerville, C. (2007) The *Arabidopsis irregular xylem8* mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. *Plant Cell*, **19**, 237–255.
133. Zhou, G.-K., Zhong, R., Richardson, E.A., Morrison III, W.H., Nairn, C.J., Wood-Jones, A. & Ye, Z.-H. (2006) The poplar glycosyltransferase GT47C is functionally conserved with Arabidopsis Fragile fiber8. *Plant and Cell Physiology*, **47**, 1229–1240.
134. Watt, G., Leoff, C., Harper, A.D. & Bar-Peled, M. (2004) A bifunctional 3,5-epimerase/4-keto reductase for nucleotide-rhamnose synthesis in Arabidopsis. *Plant Physiology*, **134**, 1337–1346.
135. Bouton, S., Leboeuf, E., Mouille, G., Leydecker, M.T., Talbotec, J., Granier, F., Lahaye, M., Höfte, H. & Truong, H.-N. (2002) *QUASIMODO1* encodes a putative membrane-bound glycosyltransferase required for normal pectin synthesis and cell adhesion in Arabidopsis. *Plant Cell*, **14**, 2577–2590.
136. Lao, N.T., Long, D., Kiang, S., Coupland, G., Shoue, D.A., Carpita, N.C. & Kavanagh, T.A. (2003) Mutation of a family 8 glycosyltransferase gene alters cell wall carbohydrate composition and causes a humidity-sensitive semi-sterile dwarf phenotype in *Arabidopsis*. *Plant Molecular Biology*, **53**, 687–701.
137. Sterling, J.D., Atmodjo, M.A., Inwood, S.E., Kolli, V.S.K., Quigley, H.F., Hahn, M.G. & Mohnen, D. (2006) Functional identification of an *Arabidopsis* pectin biosynthetic homogalacturonan galacturonosyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 5236–5241.
138. Coutinho, P.M., Deleury, E., Davies, G.J. & Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology*, **328**, 307–317.
139. Aspeborg, H., Schrader, J., Coutinho, P.M., Stam, M., Kallas, Å., Djerbi, S., Nilsson, P., Denman, S., Amini, B., Sterky, F., Master, E., Sandberg, G., Mellerowicz, E., Sundberg, B., Henrissat, B. & Teeri, T.T. (2005) Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. *Plant Physiology*, **137**, 983–997.
140. Wu, Y.T. & Liu, J.Y. (2005) Molecular cloning and characterization of a cotton glucuronosyltransferase gene. *Journal of Plant Physiology*, **162**, 573–582.
141. Baydoun, E.A.-H. & Brett, C.T. (1997) Distribution of xylosyltransferases and glucuronyltransferase within the Golgi apparatus in etiolated pea (*Pisum sativum* L.) epicotyls. *Journal of Experimental Botany*, **48**, 1209–1214.
142. Gregory, A.C.E., Smith, C., Kerry, M.E., Wheatley, E.R. & Bolwell, G.P. (2002) Comparative subcellular immunolocalization of polypeptides associated with xylan and callose synthases in French bean (*Phaseolus vulgaris*) during secondary wall formation. *Phytochemistry*, **59**, 249–259.
143. Carnachan, S.M. & Harris, P.J. (2000) Ferulic acid is bound to the primary cell walls of all gymnosperm families. *Biochemical Systematics and Ecology*, **28**, 865–879.

144. Harris, P.J., Kelderman, M.R., Kendon, M.F. & McKenzie, R.J. (1997) Monosaccharide composition of unligified cell walls of monocotyledons in relation to the occurrence of wall-bound ferulic acid. *Biochemical Systematics and Ecology*, **25**, 167–179.
145. Smith, B.G. & Harris, P.J. (1999) The polysaccharide composition of Poales cell walls: Poaceae are not unique. *Biochemical Systematics and Ecology*, **27**, 33–53.
146. Harris, P.J. (2005) Diversity in plant cell walls. In: *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants* (ed. R.J. Henry), pp. 201–227. CAB International Publishing, Wallingford, Oxon, UK.
147. Popper, Z.A. & Fry, S.C. (2004) Primary cell wall composition of pteridophytes and spermatophytes. *New Phytologist*, **164**, 165–174.
148. Popper, Z.A. & Fry, S.C. (2003) Primary cell wall composition of bryophytes and charophytes. *Annals of Botany*, **91**, 1–12.
149. Staudte, R.G., Woodward, J.R., Fincher, G.B. & Stone, B.A. (1983) Water-soluble (1,3),(1,4)- β -D-glucans from barley (*Hordeum vulgare*) endosperm: III. Distribution of cellotriosyl and cellotetraosyl residues. *Carbohydrate Polymer*, **3**, 299–312.
150. Gibeaut, D.M. & Carpita, N.C. (1993) Glucan synthesis in membranes from Zea mays and Glycine max: Interaction of ER and Golgi membranes. *Plant Physiology*, **102**, 51–51.
151. Urbanowicz, B.R., Rayon, C. & Carpita, N.C. (2004) Topology of the maize mixed linkage (1 \rightarrow 3), (1 \rightarrow 4)- β -D-glucan synthase at the Golgi membrane. *Plant Physiology*, **134**, 758–768.
152. Buckeridge, M.S., Vergara, C.E. & Carpita, N.C. (1999) The mechanism of synthesis of a mixed-linkage (1 \rightarrow 3),(1 \rightarrow 4) β -D-glucan in maize. Evidence for multiple sites of glucosyl transfer in the synthase complex. *Plant Physiology*, **120**, 1105–1116.
153. Vergara, C.E. & Carpita, N.C. (2001) β -D-Glycan synthases and the CesA gene family: Lessons to be learned from the mixed-linkage (1 \rightarrow 3);(1 \rightarrow 4) β -D-glucan synthase. *Plant Molecular Biology*, **47**, 145–160.
154. Burton, R.A., Wilson, S.M., Hrmova, M., Harvey, A.J., Shirley, N.J., Medhurst, A., Stone, B.A., Newbigin, E.J., Bacic, A. & Fincher, G.B. (2006) Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)- β -D-glucans. *Science*, **311**, 1940–1942.
155. Gibeaut, D.M. (2000) Nucleotide sugars and glycosyltransferases for synthesis of cell wall matrix polysaccharides. *Plant Physiology and Biochemistry*, **38**, 69–80.
156. Gibeaut, D.M. & Carpita, N.C. (1991) Tracing cell wall biogenesis in intact cells and plants selective turnover and alteration of soluble and cell wall polysaccharides in grasses. *Plant Physiology*, **97**, 551–561.
157. O'Neill, M., Albersheim, P. & Darvill, A. (1990) The pectic polysaccharides of primary cell walls. In: *Methods in Plant Biochemistry*, Vol. 2 (ed. P.M. Dey), pp. 415–441. Academic Press, London.
158. Ridley, B.L., O'Neill, M.A. & Mohnen, D. (2001) Pectins: Structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*, **57**, 929–967.
159. O'Neill, M.A., Ishii, T., Albersheim, P. & Darvill, A.G. (2004) Rhamnogalacturonan II: Structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annual Review of Plant Biology*, **55**, 109–139.
160. Willats, W.G.T., McCartney, L., Mackie, W. & Knox, J.P. (2001) Pectin: Cell biology and prospects for functional analysis. *Plant Molecular Biology*, **47**, 9–27.
161. O'Neill, M.A., Eberhard, S., Albersheim, P. & Darvill, A.G. (2001) Requirement of borate cross-linking of cell wall rhamnogalacturonan II for *Arabidopsis* growth. *Science*, 846–849.
162. Fleischer, A., O'Neill, M.A. & Ehwald, R. (1999) The pore size of non-graminaceous plant cell walls is rapidly decreased by borate ester cross-linking of the pectic polysaccharide rhamnogalacturonan II. *Plant Physiology*, **121**, 829–838.
163. Shomer, I., Novacky, A.J., Pike, S.M., Yermiyahu, U. & Kinraide, T.B. (2003) Electrical potentials of plant cell walls in response to the ionic environment. *Plant Physiology*, **133**, 411–422.

164. Atkinson, R.G., Schröder, R., Hallett, I.C., Cohen, D. & MacRae, E.A. (2002) Overexpression of polygalacturonase in transgenic apple trees leads to a range of novel phenotypes involving changes in cell adhesion. *Plant Physiology*, **129**, 122–133.
165. Rhee, S.Y., Osborne, E., Poindexter, P.D. & Somerville, C.R. (2003) Microspore separation in the *quartet* 3 mutants of *Arabidopsis* is impaired by a defect in a developmentally regulated polygalacturonase required for pollen mother cell wall degradation. *Plant Physiology*, **133**, 1170–1180.
166. Mohnen, D. & Hahn, M.G. (1993) Cell wall carbohydrates as signals in plants. *Seminars in Cell Biology*, **4**, 93–102.
167. Ferrari, S., Galletti, R., Denoux, C., De Lorenzo, G., Ausubel, F.M. & Dewdney, J. (2007) Resistance to *Botrytis cinerea* induced in *Arabidopsis thaliana* by elicitors is independent of salicylic acid, ethylene, or jasmonate signaling but requires *PAD3*. *Plant Physiology*, **144**, 367–379.
168. Mollet, J.-C., Park, S.-Y., Nothnagel, E.A. & Lord, E.M. (2000) A lily stylar pectin is necessary for pollen tube adhesion to an in vitro stylar matrix. *Plant Cell*, **12**, 1737–1749.
169. Western, T.L., Burn, J., Tan, W.L., Skinner, D.J., Martin-McCaffrey, L., Moffatt, B.A. & Haughn, G.W. (2001) Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in *Arabidopsis*. *Plant Physiology*, **127**, 998–1011.
170. Willats, W.G.T., McCartney, L. & Knox, J.P. (2001) In-situ analysis of pectic polysaccharides in seed mucilage and at the root surface of *Arabidopsis thaliana*. *Planta*, **213**, 37–44.
171. Usadel, B., Kuschinsky, A.M., Rosso, M.G., Eckermann, N. & Pauly, M. (2004) RHM2 is involved in mucilage pectin synthesis and is required for the development of the seed coat in *Arabidopsis*. *Plant Physiology*, **134**, 286–295.
172. González-Carranza, Z.H., Whitelaw, C.A., Swarup, R. & Roberts, J.A. (2002) Temporal and spatial expression of a polygalacturonase during leaf and flower abscission in oilseed rape and *Arabidopsis*. *Plant Physiology*, **128**, 534–543.
173. Jones, L., Milne, J.L., Ashford, D. & Mason, S.J.M. (2003) Cell wall arabinan is essential for guard cell function. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 11783–11788.
174. Marfà, V., Gollin, D.J., Eberhard, S., Mohnen, D., Darvill, A. & Albersheim, P. (1991) Oligogalacturonides are able to induce flowers to form on tobacco explants. *Plant Journal*, **1**, 217–225.
175. Skjot, M., Pauly, M., Bush, M.S., Borkhardt, B., McCann, M.C. & Ulvskov, P. (2002) Direct interference with rhamnogalacturonan I biosynthesis in Golgi vesicles. *Plant Physiology*, **129**, 95–102.
176. Brown, K. (2001) Xylem may direct water where it's needed. *Science*, **291**, 571–572.
177. Côté, F., Ham, K.-S., Hahn, M.G. & Bergmann, C.W. (1998) Oligosaccharide elicitors in host-pathogen interactions generation, perception, and signal transduction. In: *Plant-Microbe Interactions*, Vol. 29 (eds B.B. Biswas & H.K. Das), pp. 385–432. Plenum Press, New York.
178. Fry, S., Aldington, S., Hetherington, P. & Aitken, J. (1993) Oligosaccharides as signals and substrates in the plant cell wall. *Plant Physiology*, **103**, 1–5.
179. Shibuya, N. & Minami, E. (2001) Oligosaccharide signalling for defence responses in plant. *Physiological and Molecular Plant Pathology*, **59**, 223–233.
180. Capodicasa, C., Vairo, D., Zabolina, O., McCartney, L., Caprari, C., Mattei, B., Manfredini, C., Aracri, B., Benen, J., Knox, J.P., De Lorenzo, G. & Cervone, F. (2004) Targeted modification of homogalacturonan by transgenic expression of a fungal polygalacturonase alters plant growth. *Plant Physiology*, **135**, 1294–1304.
181. Iwai, H., Masaoka, N., Ishii, T. & Satoh, S. (2002) A pectin glucuronyltransferase gene is essential for intercellular attachment in the plant meristem. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16319–16324.
182. Bouveng, H.O. (1965) Polysaccharides in pollen II. The xylogalacturonan from mountain pine (*Pinus mugo* Turra) pollen. *ACTA Chemica Scandinavica*, **19**, 953–963.

183. Schols, H.A., Posthumus, M.A. & Voragen, A.G.J. (1990) Structural features of hairy regions of pectins isolated from apple juice produced by the liquefaction process. *Carbohydrate Research*, **206**, 117–129.
184. Schols, H.A., Vierhuis, E., Bakx, E.J. & Voragen, A.G.J. (1995) Different populations of pectic hairy regions occur in apple cell walls. *Carbohydrate Research*, **275**, 343–360.
185. Yu, L. & Mort, A.J. (1996) Partial characterization of xylogalacturonans from cell walls of ripe watermelon fruit: Inhibition of endopolygalacturonase activity by xylosylation. In: *Pectins and Pectinases*, Vol. 14 (eds. J. Visser & A.G.J. Voragen), pp. 79–88. Elsevier, Amsterdam.
186. Kikuchi, A., Edashige, Y., Ishii, T. & Satoh, S. (1996) A xylogalacturonan whose level is dependent on the size of cell clusters is present in the pectin from cultured carrot cells. *Planta*, **200**, 369–372.
187. Zandleven, J., Sørensen, S.O., Harholt, J., Beldman, G., Schols, H.A., Scheller, H.V. & Voragen, A.J. (2007) Xylogalacturonan exists in cell walls from various tissues of *Arabidopsis thaliana*. *Phytochemistry*, **68**, 1219–1226.
188. Hart, D.A. & Kindel, P.K. (1970) Isolation and partial characterization of apiogalacturonans from the cell wall of *Lemna minor*. *Biochemical Journal*, **116**, 569–579.
189. Watson, R.R. & Orenstein, N.S. (1975) Chemistry and biochemistry of apiose. *Advances in Carbohydrate Chemistry and Biochemistry*, **31**, 135–184.
190. Stephen, A.M. (1983) Other plant polysaccharides. In: *The Polysaccharides*, Vol. 2 (ed. G.O. Aspinall), pp. 97–193. Academic Press, New York.
191. Cheng, L. & Kindel, P.K. (1997) Detection and homogeneity of cell wall pectic polysaccharides of *Lemna minor*. *Carbohydrate Research*, **301**, 205–212.
192. Mohnen, D. (2002) Biosynthesis of pectins. In: *Pectins and their Manipulation* (eds. G.B. Seymour & J.P. Knox), pp. 52–98. Blackwell Publishing and CRC Press, Oxford.
193. Schols, H.A., Bakx, E.J., Schipper, D. & Voragen, A.G.J. (1995) A xylogalacturonan subunit present in the modified hairy regions of apple pectin. *Carbohydrate Research*, **279**, 265–279.
194. Mohnen, D., Doong, R.L., Liljebjelke, K., Fralish, G. & Chan, J. (1996) In: *Cell Free Synthesis of the Pectic Polysaccharide Homogalacturonan, Pectins and Pectinases* (eds. J. Visser & A.G.J. Voragen), pp. 109–126. Elsevier Science B.V., Amsterdam.
195. Mort, A.J., Qiu, F. & Maness, N.O. (1993) Determination of the pattern of methyl esterification in pectin. Distribution of contiguous nonesterified residues. *Carbohydrate Research*, **247**, 21–35.
196. Rombouts, F.M. & Thibault, J.F. (1986) Sugar beet pectins: Chemical structure and gelation through oxidative coupling. In: *Chemistry and Function of Pectins* (eds. M.L. Fishman & J.J. Jen), pp. 49–60. American Chemical Society, Washington, DC.
197. De Vries, J.A., Voragen, A.G.J., Rombouts, F.M. & Pilnik, W. (1986) Structural studies of apple pectins with pectolytic enzymes. In: *Chemistry and Function of Pectins* (eds. M.L. Fishman & J.J. Jen), pp. 38–48. American Chemical Society, Washington, DC.
198. Ishii, T. (1995) Pectic polysaccharides from bamboo shoot cell-walls. *Mokuzai Gakkaishi*, **41**, 669–676.
199. Ishii, T. (1997) O-Acetylated oligosaccharides from pectins of potato tuber cell walls. *Plant Physiology*, **113**, 1265–1272.
200. Kim, J.-B. & Carpita, N.C. (1992) Changes in esterification of the uronic acid groups of cell wall polysaccharides during elongation of maize coleoptiles. *Plant Physiology*, **98**, 646–653.
201. Brown, J.A. & Fry, S.C. (1993) Novel O-D-galacturonoyl esters in the pectic polysaccharides of suspension-cultured plant cells. *Plant Physiology*, **103**, 993–999.
202. McCann, M.C., Shi, J., Roberts, K. & Carpita, N.C. (1994) Changes in pectin structure and localization during the growth of unadapted and NaCl-adapted tobacco cells. *Plant Journal*, **5**, 773–785.
203. Hou, W.-C. & Chang, W.-H. (1996) Pectinesterase-catalyzed firming effects during precooking of vegetables. *Journal of Food Biochemistry*, **20**, 397–416.
204. Djelineo, I. (2001) *Structural Studies of Pectin*. PhD thesis, The University of Georgia.

205. Mohnen, D. (1999) Biosynthesis of pectins and galactomannans. In: *Comprehensive Natural Products Chemistry, Carbohydrates and Their Derivatives Including Tannins, Cellulose, and Related Lignins*, Vol. 3 (ed. B.M. Pinto), pp. 497–527. Elsevier, Oxford.
206. O'Neill, M.A. & York, W.S. (2003) The composition and structure of plant primary cell walls. In: *The Plant Cell Wall*, Annual Plant Reviews (ed. J.K.C. Rose), pp. 1–54. Blackwell Publishing/CRC Press, Oxford, UK/Boca Raton, FL.
207. Aspinall, G.O., Begbie, R., Hamilton, A. & Whyte, J.N.C. (1967) Polysaccharides of soy-beans. Part III. Extraction and fractionation of polysaccharides from cotyledon meal. *Journal of the Chemical Society, (C)* 1065–1070.
208. Nakamura, A., Furuta, H., Maeda, H., Takao, T. & Nagamatsu, Y. (2002) Structural studies by stepwise enzymatic degradation of the main backbone of soybean soluble polysaccharides consisting of galacturonan and rhamnogalacturonan. *Bioscience, Biotechnology and Biochemistry*, **66**, 1301–1313.
209. Scheller, H.V., Jensen, J.K., Sørensen, S.O., Harholt, J. & Geshi, N. (2007) Biosynthesis of pectin. *Physiologica Plantarum*, **129**, 283–295.
210. Price, N.J., Reiter, W.-D. & Raikhel, N.V. (2001) Plant glycosyltransferases. *Current Opinion in Plant Biology*, **4**, 219–224.
211. Zhong, R. & Ye, Z.-H. (2003) Unraveling the functions of glycosyltransferase family 47 in plants. *Trends in Plant Science*, **8**, 565–568.
212. Scheible, W.-R. & Pauly, M. (2004) Glycosyltransferases and cell wall biosynthesis: Novel players and insights. *Current Opinion in Plant Biology*, **7**, 285–295.
213. Farrokhi, N., Burton, R.A., Brownfield, L., Hrmova, M., Wilson, S.M., Bacic, A. & Fincher, G.B. (2006) Plant cell wall biosynthesis: Genetic, biochemical and functional genomics approaches to the identification of key genes. *Plant Biotechnology Journal*, **4**, 145–167.
214. Egelund, J., Skjot, M., Geshi, N., Ulvskov, P. & Petersen, B.L. (2004) A complementary bioinformatics approach to identify potential plant cell wall glycosyltransferase-encoding genes. *Plant Physiology*, **136**, 2609–2620.
215. Yokoyama, R. & Nishitani, K. (2004) Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and Arabidopsis, *Plant and Cell Physiology*, **45**, 1111–1121.
216. Yong, W., Link, B., O'Malley, R., Tewari, J., Hunter, C.T., Lu, C.-A., Li, X., Bleecker, A.B., Koch, K.E., McCann, M.C., McCarty, D.R., Patterson, S.E., Reiter, W.-D., Staiger, C., Thomas, S.R., Vermerris, W. & Carpita, N.C. (2005) Genomics of plant cell wall biogenesis. *Planta*, **221**, 747–751.
217. Geisler-Lee, J., Geisler, M., Coutinho, P.M., Segerman, B., Nishikubo, N., Takahashi, J., Aspeborg, H., Djerbi, S., Master, E., Andersson-Gunneras, S., Sundberg, B., Karpinski, S., Teeri, T.T., Kleczkowski, L.A., Henrissat, B. & Mellerowicz, E.J. (2006) Poplar carbohydrate-active enzymes. Gene identification and expression analyses. *Plant Physiology*, **140**, 946–962.
218. Pilling, E. & Höfte, H. (2003) Feedback from the wall. *Current Opinion in Plant Biology*, **6**, 611–616.
219. Johansen, J.N., Vernhettes, S. & Höfte, H. (2006) The ins and outs of plant cell walls. *Current Opinion in Plant Biology*, **9**, 616–620.
220. Ko, J.-H., Beers, E.P. & Han, K.-H. (2006) Global comparative transcriptome analysis identifies gene network regulating secondary xylem development in *Arabidopsis thaliana*. *Molecular Genetics and Genomics*, **276**, 517–531.
221. Northcote, D.H. & Pickett-Heaps, J.D. (1966) A function of the Golgi apparatus in polysaccharide synthesis and transport in the root-cap cells of wheat. *Biochemical Journal*, **98**, 159–167.
222. Northcote, D.H. (1971) The Golgi apparatus. *Endeavor*, **30**, 26–33.
223. Moore, P.J., Swords, K.M., Lynch, M.A. & Staehelin, L.A. (1991) Spatial organization of the assembly pathways of glycoproteins and complex polysaccharides in the Golgi apparatus of plants. *Journal of Cell Biology*, **112**, 589–602.

224. Staehelin, L.A. & Moore, I. (1995) The plant Golgi apparatus: Structure, functional organization and trafficking mechanisms. *Annual Review of Plant Physiology and Plant Molecular Biology*, **46**, 261–288.
225. Willats, W.G.T., Steele-King, C.G., McCartney, L., Orfila, C., Marcus, S.E. & Knox, J.P. (2000) Making and using antibody probes to study plant cell walls. *Plant Physiology and Biochemistry*, **38**, 27–36.
226. Vannier, M.P., Thoiron, B., Morvan, C. & Demarty, M. (1992) Localization of methyltransferase activities throughout the endomembrane complex system of flax (*Linum usitatissimum* L.) hypocotyls. *Biochemical Journal*, **286**, 863–868.
227. Burlard, T., Schaumann-Gaudinet, A., Bruyant-Vannier, M.-P. & Morvan, C. (1997) Various pectin methyltransferase activities with affinity for low and highly methylated pectins. *Plant and Cell Physiology*, **38**, 259–267.
228. Goubet, F. & Mohnen, D. (1999) Subcellular localization and topology of homogalacturonan methyltransferase in suspension-cultured *Nicotiano tabacum* cells. *Planta*, **209**, 112–117.
229. Sterling, J.D., Quigley, H.F., Orellana, A. & Mohnen, D. (2001) The catalytic site of the pectin biosynthetic enzyme α -1,4-galacturonosyltransferase is located in the lumen of the Golgi. *Plant Physiology*, **127**, 360–371.
230. Baydoun, E.A.H., Abdel-Massih, R.M., Dani, D., Rizk, S.E. & Brett, C.T. (2001) Galactosyl- and fucosyltrasferases in etiolated pea epicotyls: Product identification and sub-cellular localisation. *Journal of Plant Physiology*, **158**, 145–150.
231. Höfte, H. & Staehelin, L.A. (2000) Cell Biology Plant cells do it differently. *Current Opinions in Plant Biology* 2000, **3**, 447–449.
232. Nebenführ, A., Gallagher, L.A., Dunahay, T.G., Frohlick, J.A., Mazurkiewicz, A.M., Meehl, J.B. & Staehelin, L.A. (1999) Stop-and-go movements of plant Golgi stacks are mediated by the acto-myosin system. *Plant Physiology*, **121**, 1127–1141.
233. Lynch, M.A. & Staehelin, L.A. (1992) Domain-specific and cell type-specific localization of two types of cell wall matrix polysaccharides in the clover root tip. *Journal of Cell Biology*, **118**, 467–479.
234. Zhang, G.F. & Staehelin, L.A. (1992) Functional compartmentation of the Golgi apparatus of plant cells; Immunocytochemical analysis of high-pressure frozen- and freeze-substituted sycamore maple suspension culture cells. *Plant Physiology*, **99**, 1070–1083.
235. VandenBosch, K.A., Bradley, D.J., Knox, J.P., Perotto, S., Butcher, G.W. & Brewin, N.J. (1989) Common components of the infection thread matrix and the intercellular space identified by immunocytochemical analysis of pea nodules and uninfected roots. *EMBO Journal*, **8**, 335–341.
236. Knox, J.P., Linstead, P.J., King, J., Cooper, C. & Roberts, K. (1990) Pectin esterification is spatially regulated both within cell walls and between developing tissues of root apices. *Planta*, **181**, 512–521.
237. Moore, P.J., Darvill, A.G., Albersheim, P. & Staehelin, L.A. (1986) Immunogold localization of xyloglucan and rhamnogalacturonan I in the cell walls of suspension-cultured sycamore cells. *Plant Physiology*, **82**, 787–794.
238. Liners, F., Letesson, J.-J., Didembourg, C. & Van Cutsem, P. (1989) Monoclonal antibodies against pectin: Recognition of a conformation induced by calcium. *Plant Physiology*, **91**, 1419–1424.
239. Vian, B. & Roland, J.-C. (1991) Affinodetection of the sites of formation and of the further distribution of polygalacturonans and native cellulose in growing plant cells. *Biology of the Cell*, **71**, 43–55.
240. Liners, F. & Van Cutsem, P. (1992) Distribution of pectic polysaccharides throughout walls of suspension-cultured carrot cells. *Protoplasma*, **170**, 10–21.
241. Sherrier, D.J. & VandenBosch, K.A. (1994) Secretion of cell wall polysaccharides in *Vicia* root hairs. *Plant Journal* **5**, 185–195.

242. Carpita, N.C. & Gibeaut, D.M. (1993) Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth. *Plant Journal*, **3**, 1–30.
243. Liners, F., Gaspar, T. & Van Cutsem, P. (1994) Acetyl- and methyl-esterification of pectins of friable and compact sugar-beet calli: Consequences for intercellular adhesion. *Planta*, **192**, 545–556.
244. Casero, P.J. & Knox, J.P. (1995) The monoclonal antibody JIM5 indicates patterns of pectin deposition in relation to pit fields at the plasma-membrane-face of tomato pericarp cell walls. *Protoplasma*, **188**, 133–137.
245. Dolan, L., Linstead, P. & Roberts, K. (1997) Developmental regulation of pectic polysaccharides in the root meristem of *Arabidopsis*. *Journal of Experimental Botany*, **48**, 713–720.
246. Micheli, F. (2001) Pectin methylesterases: Cell wall enzymes with important roles in plant physiology. *Trends in Plant Science*, **6**, 414–419.
247. Stoddart, R.W. & Northcote, D.H. (1967) Metabolic relationships of the isolated fractions of the pectic substances of actively growing sycamore cells. *Biochemical Journal*, **105**, 45–59.
248. Shea, E.M., Gibeaut, D.M. & Carpita, N.C. (1989) Structural analysis of the cell walls regenerated by carrot protoplasts. *Planta*, **179**, 293–308.
249. Li, Y.Q., Chen, F., Linskens, H.F. & Cresti, M. (1994) Distribution of unesterified and esterified pectins in cell walls of pollen tubes of flowering plants. *Sex Plant Reproduction*, **7**, 145–152.
250. Marty, P., Goldberg, R., Liberman, M., Vian, B., Bertheau, Y. & Jouan, B. (1995) Composition and localization of pectic polymers in the stems of two *Solanum tuberosum* genotypes. *Plant Physiology and Biochemistry*, **33**, 409–417.
251. Fujiki, Y., Hubbard, A.L., Fowler, S. & Lazarow, P.B. (1982) Isolation of intracellular membranes by means of sodium carbonate treatment: Application to endoplasmic reticulum. *Journal of Cell Biology*, **93**, 97–102.
252. Stacey, N.J., Roberts, K., Carpita, N.C., Wells, B. & McCann, M.C. (1995) Dynamic changes in cell surface molecules are very early events in the differentiation of mesophyll cells from *Zinnia elegans* into tracheary elements. *Plant Journal*, **8**, 891–906.
253. Willats, W.G., Steele-King, C.G., Marcus, S.E. & Knox, J.P. (1999) Side chains of pectin polysaccharides are regulated in relation to cell proliferation and cell differentiation. *Plant Journal*, **20**, 619–628.
254. Orfila, C. & Knox, J.P. (2000) Spatial regulation of pectic polysaccharides in relation to pit fields in cell walls of tomato fruit pericarp. *Plant Physiology*, **122**, 775–781.
255. McCartney, L., Ormerod, A.P., Gidley, M.J. & Knox, J.P. (2000) Temporal and spatial regulation of pectic (1-4)-beta-d-galactan in cell walls of developing pea cotyledons: Implications for mechanical properties. *Plant Journal*, **22**, 105–113.
256. Whitelock, J.M. & Iozzo, R.V. (2005) Heparan sulfate: A complex polymer charged with biological activity. *Chemical Review*, **105**, 2745–2764.
257. Crosthwaite, S.K., Macdonald, F.M., Baydoun, E.A.H. & Brett, C.T. (1994) Properties of a protein-linked glucuronoxylan formed in the plant Golgi-apparatus. *Journal of Experimental Botany*, **45**, 471–475.
258. Abdel-Massih, R.M., Baydoun, E.A. & Brett, C.T. (2003) *In vitro* biosynthesis of 1,4-β-galactan attached to a pectin-xyloglucan complex in pea. *Planta*, **216**, 502–511.
259. Cumming, C.M., Rizkallah, H.D., McKendrick, K.A., Abdel-Massih, R.M., Baydoun, E.A. & Brett, C.T. (2005) Biosynthesis and cell-wall deposition of a pectin-xyloglucan complex in pea. *Planta*, **222**, 546–555.
260. Orfila, C., Sørensen, S.O., Harholt, J., Geshi, N., Crombie, H., Truong, H.-N., Reid, J.S., Knox, J.P. & Scheller, H.V. (2005) *QUASIMODO1* is expressed in vascular tissue of *Arabidopsis thaliana* inflorescence stems, and affects homogalacturonan and xylan biosynthesis. *Planta*, **222**, 613–622.

261. Lau, J.M., McNeil, M., Darvill, A.G. & Albersheim, P. (1985) Structure of the backbone of rhamnogalacturonan I, a pectic polysaccharide in the primary cell walls of plants. *Carbohydrate Research*, **137**, 111–125.
262. Eda, S., Miyabe, K., Akiyama, Y., Ohnishi, A. & Kato, K. (1986) A pectic polysaccharide from cell walls of tobacco (*Nicotiana tabacum*) mesophyll. *Carbohydrate Research*, **158**, 205–216.
263. Nakamura, A., Furuta, H., Maeda, H., Takao, T. & Nagamatsu, Y. (2002) Analysis of the molecular construction of xylogalacturonan isolated from soluble soybean polysaccharides. *Bioscience, Biotechnology and Biochemistry*, **66**, 1155–1158.
264. Lau, J.M., McNeil, M., Darvill, A.G. & Albersheim, P. (1987) Treatment of rhamnogalacturonan I with lithium in ethylenediamine. *Carbohydrate Research*, **168**, 245–274.
265. Morita, M. (1965) Polysaccharides of soybean seeds. Part I. Polysaccharide constituents of “hot-water-extract” fractions of soybean seeds and an arabinogalactan as its major component. *Agricultural and Biological Chemistry*, **29**, 564–573.
266. Morita, M. (1965) Polysaccharides of soybean seeds. Part II. A methylated arabinogalactan isolated from methylated product of “hot-water extract” fraction of soybean seed polysaccharides. *Agricultural and Biological Chemistry*, **29**, 626–630.
267. Aspinall, G.O. (1980) Chemistry of cell wall polysaccharides. In: *The Biochemistry of Plants*, Vol. 3 (ed. J. Preiss), pp. 473–500. Academic Press, New York.
268. Huisman, M.M.H., Brüll, L.P., Thomas-Oates, J.E., Haverkamp, J., Schols, H.A. & Voragen, A.G.J. (2001) The occurrence of internal (1-5)-linked arabinofuranose and arabinopyranose residues in arabinogalactan side chains from soybean pectic substances. *Carbohydrate Research*, **330**, 103–114.
269. O'Neill, M.A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A.G. & Albersheim, P. (1997) Rhamnogalacturonan-II, a pectic polysaccharide in the walls of growing plant cell, forms a dimer that is covalently cross-linked by a borate ester – in vitro conditions for the formation and hydrolysis of the dimer. *Journal of Biological Chemistry*, **272**, 3869.
270. Harholt, J., Jensen, J.K., Sorensen, S.O., Orfila, C., Pauly, M. & Scheller, H.V. (2006) ARABINAN DEFICIENT 1 is a putative arabinosyltransferase involved in biosynthesis of pectic arabinan in arabidopsis. *Plant Physiology*, **140**, 49–58.
271. Ishii, T., Konishi, T., Ito, Y., Ono, H., Ohnishi-Kameyama, M. & Maeda, I. (2005) A β -(1 \rightarrow 3)-arabinopyranosyltransferase that transfers a single arabinopyranose onto arabinooligosaccharides in mung bean (*Vigna radiata*) hypocotyls. *Phytochemistry*, **66**, 2418–2425.
272. Odonmažig, P., Ebringerová, A., Machová, E. & Alföldi, J. (1994) Structural and molecular properties of the arabinogalactan isolated from Mongolian larchwood (*Larix dahurica* L.). *Carbohydrate Research*, **252**, 317–324.
273. Ishii, T., Ono, H., Ohnishi-Kameyama, M. & Maeda, I. (2005) Enzymic transfer of α -L-arabinopyranosyl residues to exogenous 1,4-linked β -D-galacto-oligosaccharides using solubilized mung bean (*Vigna radiata*) hypocotyl microsomes and UDP- β -L-arabinopyranose. *Planta*, **221**, 953–963.
274. Egelund, J., Petersen, B.L., Motawia, M.S., Damager, I., Faik, A., Olsen, C.E., Ishii, T., Clausen, H., Ulvskov, P. & Geshi, N. (2006) *Arabidopsis thaliana* RGXT1 and RGXT2 encode Golgi-localized (1,3)- α -D-xylosyltransferases involved in the synthesis of pectic rhamnogalacturonan II. *Plant Cell*, **18**, 2593–2607.
275. Yamada, H., Yanahira, S., Kiyohara, H., Cyong, J.-C. & Otsuka, Y. (1987) Characterization of anti-complementary acidic heteroglycans from the seed of *Coix lacryma-jobi* var. *ma-yuen*. *Phytochemistry*, **26**, 3269–3275.
276. Duan, J., Wang, X., Dong, Q., Fang, J. & Li, X. (2003) Structural features of a pectic arabinogalactan with immunological activity from the leaves of *Diospyros kaki*. *Carbohydrate Research*, **338**, 1291–1297.

277. Duan, J., Zheng, Y., Dong, Q. & Fang, J. (2004) Structural analysis of a pectic polysaccharide from the leaves of *Diospyros kaki*. *Phytochemistry*, **65**, 609–615.
278. An, J., O'Neill, M.A., Albersheim, P. & Darvill, A.G. (1994) Isolation and structural characterization of β -D-glucosyluronic acid and 4-O-methyl β -D-glucosyluronic acid-containing oligosaccharides from the cell-wall pectic polysaccharide, rhamnogalacturonan I. *Carbohydrate Research*, **252**, 235–243.
279. Gaspar, Y., Johnson, K.L., McKenna, J.A., Bacic, A. & Schultz, C.J. (2001) The complex structures of arabinogalactan-proteins and the journey towards understanding function. *Plant Molecular Biology*, **47**, 161–176.
280. Kauss, H., Swanson, A.L., Arnold, R. & Odzuck, W. (1969) Biosynthesis of pectic substances. Localization of enzymes and products in a lipid-membrane complex. *Biochimica Biophysica Acta*, **91**, 55–61.
281. Kauss, H., Swanson, A.L. & Hassid, W.Z. (1967) Biosynthesis of the methyl ester groups of pectin by transmethylation from S-adenosyl-L-methionine. *Biochemical and Biophysical Research Communications*, **26**, 234–240.
282. Schaumann, A., Bruyant-Vannier, M.-P., Goubet, F. & Morvan, C. (1993) Pectic metabolism in suspension-cultured cells of flax, *Linum usitatissimum*. *Plant and Cell Physiology*, **34**, 891–897.
283. Crombie, H.J. & Reid, J.S.G. (1998) Pectin methyltransferase: Activities in particulate and solubilised preparations from mung bean (*Vigna radiata*) hypocotyls and tomato (*Lycopersicon esculentum*) pericarp [Abstract no. 1.39]. In: *Cell Walls '98 – 8th International Cell Walls Meeting*, John Innes Centre, Norwich, UK, 1st–5th September 1998.
284. Goubet, F., Council, L.N. & Mohnen, D. (1998) Identification and partial characterization of the pectin methyltransferase “homogalacturonan-methyltransferase” from membranes of tobacco cell suspensions. *Plant Physiology*, **116**, 337–347.
285. Ishikawa, M., Kuroyama, H., Takeuchi, Y. & Tsumuraya, Y. (2000) Characterization of pectin methyltransferase from soybean hypocotyls. *Planta*, **210**, 782–791.
286. Kunkel, A., Gunter, S. & Watzig, H. (1998) Capillary electrophoresis for the quantitation of insulin in pharmaceutical quality control: Strategies to obtain maximum precision and comparison to HPLC. *American Laboratory*, **30**, 76c–89c.
287. Mouille, G., Ralet, M.C., Cavelier, C., Eland, C., Effroy, D., Hematy, K., McCartney, L., Truong, H.N., Gaudon, V., Thibault, J.F., Marchant, A. & Hofte, H. (2007) Homogalacturonan synthesis in *Arabidopsis thaliana* requires a Golgi-localized protein with a putative methyltransferase domain. *Plant Journal*, **50**, 605–614.
288. Bacic, A., Harris, P.J. & Stone, B.A. (1988) Structure and function of plant cell walls. In: *The Biochemistry of Plants*, Vol. 14 (ed. J. Preiss), pp. 297–371. Academic Press, New York.
289. Komalavilas, P. & Mort, A.J. (1989) The acetylation at O-3 of galacturonic acid in the rhamnose-rich portion of pectins. *Carbohydrate Research*, **189**, 261–272.
290. Lerouge, P., O'Neill, M.A., Darvill, A.G. & Albersheim, P. (1993) Structural characterization of endo-glycanase-generated oligoglycosyl side chains of rhamnogalacturonan I. *Carbohydrate Research*, **243**, 359–371.
291. Ishii, T. (1997) Structure and functions of feruloylated polysaccharides. *Plant Science*, **127**, 111–127.
292. Fry, S.C. (1982) Phenolic components of the primary cell wall. Feruloylated disaccharides of D-galactose and L-arabinose from spinach polysaccharide. *Biochemical Journal*, **203**, 493–504.
293. Ralph, J., Grabber, J.H. & Hatfield, R.D. (1995) Lignin-ferulate cross-links in grasses: Active incorporation of ferulate polysaccharide esters into ryegrass lignins. *Carbohydrate Research*, **275**, 167–178.
294. Encina, A. & Fry, J.C. (2005) Oxidative coupling of a feruloyl-arabinoxylan trisaccharide (FAXX) in the walls of living maize cells requires endogenous hydrogen peroxide and is controlled by a low-Mr apoplastic inhibitor. *Planta*, **223**, 77–89.

295. Harris, P.J. & Hartley, R.D. (1980) Phenolic constituents of the cell walls of monocotyledons. *Biochemical Systematics and Ecology*, **8**, 153–160.
296. Oosterveld, A., Beldman, G., Schols, H.A. & Voragen, A.G.J. (2000) Characterization of arabinose and ferulic acid rich pectic polysaccharides and hemicelluloses from sugar beet pulp. *Carbohydrate Research*, **328**, 185–197.
297. Baydoun, E.A., Pavlencheva, N., Cumming, C.M., Waldron, K.W. & Brett, C.T. (2004) Control of dehydridiferulate cross-linking in pectins from sugar-beet tissues. *Phytochemistry*, **65**, 1107–1115.
298. Ralet, M.-C., André-Leroux, G., Quémener, B. & Thibault, J.-F. (2005) Sugar beet (*Beta vulgaris*) pectins are covalently cross-linked through diferulic bridges in the cell wall. *Phytochemistry*, **66**, 2800–2814.
299. Levigne, S.V., Ralet, M.-C.J., Quémener, B.C., Pollet, B.N., Lapierre, C. & Thibault, J.-F. (2004) Isolation from sugar beet cell walls of arabinan oligosaccharides esterified by two ferulic acid monomers. *Plant Physiology*, **134**, 1173–1180.
300. Yoshida-Shimokawa, T., Yoshida, S., Kakegawa, K. & Ishii, T. (2001) Enzymic feruloylation of arabinoxylan-trisaccharide by feruloyl-CoA: Arabinoxylan-trisaccharide O-hydroxycinnamoyl transferase from *Oryza sativa*. *Planta*, **212**, 470–474.
301. Obel, N., Porchia, A.C. & Scheller, H.V. (2003) Intracellular feruloylation of arabinoxylan in wheat: Evidence for feruloyl-glucose as precursor. *Planta*, **216**, 620–629.
302. Meyer, K., Kohler, A. & Kauss, H. (1991) Biosynthesis of ferulic acid esters of plant cell wall polysaccharides in endomembranes from parsley cells. *FEBS Letters*, **290**, 209–212.
303. Mitchell, R.A.C., Dupree, P. & Shewry, P.R. (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiology*, **144**, 43–53.
304. Thibault, J.-F., Renard, C.M.G.C., Axelos, M.A.V., Roger, P. & Crepeau, M.-J. (1993) Studies of the length of homogalacturonic regions in pectins by acid hydrolysis. *Carbohydrate Research*, **238**, 271–286.
305. Villemetz, C.L., Lin, T.-Y. & Hassid, W.Z. (1965) Biosynthesis of the polygalacturonic acid chain of pectin by a particulate enzyme preparation from *Phaseolus aureus* seedlings. *Proceedings of the National Academy of Sciences of the United States of America*, **54**, 1626–1632.
306. Lin, T.-Y., Elbein, A.D. & Su, J.C. (1966) Substrate specificity in pectin synthesis. *Biochemical and Biophysical Research Communications*, **22**, 650–657.
307. Villemetz, C.L., Swanson, A.L. & Hassid, W.Z. (1966) Properties of a polygalacturonic acid-synthesizing enzyme system from *Phaseolus aureus* seedlings. *Archives of Biochemistry and Biophysics*, **116**, 446–452.
308. Kauss, H. & Swanson, A.L. (1969) Cooperation of enzymes responsible for polymerization and methylation in pectin biosynthesis. *Zeitschrift Fur Naturforschung*, **24**, 28–33.
309. Bolwell, G.P., Dalessandro, G. & Northcote, D.H. (1985) Decrease of polygalacturonic acid synthase during xylem differentiation in sycamore. *Phytochemistry*, **24**, 699–702.
310. Doong, R.L., Liljebjelke, K., Fralish, G., Kumar, A. & Mohnen, D. (1995) Cell free synthesis of pectin: Identification and partial characterization of polygalacturonate 4- α -galacturonosyltransferase and its products from membrane preparations of tobacco (*Nicotiana tabacum* L. cv samsun) cell suspension cultures. *Plant Physiology*, **109**, 141–152.
311. Doong, R.L. & Mohnen, D. (1998) Solubilization and characterization of a galacturonosyltransferase that synthesizes the pectic polysaccharide homogalacturonan. *Plant Journal*, **13**, 363–374.
312. Scheller, H.V., Doong, R.L., Ridley, B.L. & Mohnen, D. (1999) Pectin biosynthesis: A solubilized galacturonosyltransferase from tobacco catalyzes the transfer of galacturonic acid from UDP-galacturonic acid onto the non-reducing end of homogalacturonan. *Planta*, **207**, 512–517.
313. Cumming, C.M. & Brett, C.T. (1986) A galacturonyltransferase involved in pectin biosynthesis. In: *Cell Walls '86. Proceedings of the Fourth Cell Wall Meeting* (eds. B. Vian, D. Reis & R. Goldberg), pp. 360–363, September 10–12, 1986. Université Pierre et Marie Curie – Ecole Normale Supérieure, Paris.

314. Takeuchi, Y. & Tsumuraya, Y. (2001) In vitro biosynthesis of homogalacturonan by a membrane-bound galacturonosyltransferase from epicotyls of azuki bean. *Bioscience, Biotechnology and Biochemistry*, **65**, 1519–1527.
315. Akita, K., Ishimizu, T., Tsukamoto, T., Ando, T. & Hase, S. (2002) Successive glycosyltransfer activity and enzymatic characterization of pectic polygalacturonate 4- α -galacturonosyltransferase solubilized from pollen tubes of *Petunia axillaris* using pyridylaminated oligogalacturonates as substrates. *Plant Physiology*, **130**, 374–379.
316. Guillaumie, F., Sterling, J.D., Jensen, K.J., Thomas, O.R. & Mohnen, D. (2003) Solid-supported enzymatic synthesis of pectic oligogalacturonides and their analysis by MALDI-TOF mass spectrometry. *Carbohydrate Research*, **338**, 1951–1960.
317. Liljebjelke, K., Adolphson, R., Baker, K., Doong, R.L. & Mohnen, D. (1995) Enzymatic synthesis and purification of uridine diphosphate [14 C]galacturonic acid: A substrate for pectin biosynthesis. *Analytical Biochemistry*, **225**, 296–304.
318. Orellana, A. & Mohnen, D. (1999) Enzymatic synthesis and purification of [3 H] uridine diphosphate galacturonic acid for use in studying Golgi-localized transporters. *Analytical Biochemistry*, **272**, 224–231.
319. Ishii, T. (2002) A sensitive and rapid bioassay of homogalacturonan synthase using 2-aminobenzamide-labeled oligogalacturonides. *Plant and Cell Physiology*, **43**, 1386–1389.
320. Denyer, K., Waite, D., Edwards, A., Martin, C. & Smith, A.M. (1999) Interaction with amylopectin influences the ability of granule-bound starch synthase I to elongate malto-oligosaccharides. *Biochemistry Journal*, **342**, 647–653.
321. Sterling, J.D., Lemons, J.A., Forkner, I.F. & Mohnen, D. (2005) Development of a filter assay for measuring homogalacturonan: α 1,4-galacturonosyltransferase activity. *Analytical Biochemistry*, **343**, 231–236.
322. Crombie, H.J. & Reid, J.S.G. (2001) A homogalacturonan synthase from mung bean hypocotyls [Abstract no. 131]. In: *Cell Wall '01 – 9th International Cell Wall Meeting*, Toulouse, France, September 2–7, 2001.
323. Dunkley, T.P.J., Watson, R., Griffin, J.L., Dupree, P. & Lilley, K.S. (2004) Localization of organelle proteins by isotope tagging (LOPIT). *Molecular and Cellular Proteomics*, **3**, 1128–1134.
324. Dunkley, T.P.J., Hester, S., Shadforth, I.P., Runions, J., Weimar, T., Hanton, S.L., Griffin, J.L., Bessant, C., Brandizzi, F., Hawes, C., Watson, R.B., Dupree, P. & Lilley, K.S. (2006) Mapping the *Arabidopsis* organelle proteome. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 6518–6523.
325. Leboeuf, E., Guillon, F., Thoirion, S. & Lahaye, M. (2005) Biochemical and immunohistochemical analysis of pectic polysaccharides in the cell walls of *Arabidopsis* mutant QUASIMODO 1 suspension-cultured cells: Implications for cell adhesion. *Journal of Experimental Botany*, **56**, 3171–3182.
326. Shao, M., Zheng, H., Hu, Y., Liu, D., Jang, J.-C., Ma, H. & Huang, H. (2004) The GAO-LAOZHUANGREN1 gene encodes a putative glycosyltransferase that is critical for normal development and carbohydrate metabolism. *Plant and Cell Physiology*, **45**, 1453–1460.
327. York, W.S., Darvill, A.G., McNeil, M., Stevenson, T.T. & Albersheim, P. (1985) Isolation and characterization of plant cell walls and cell wall components. *Methods in Enzymology*, **118**, 3–40.
328. Shedletsky, E., Shmuel, M., Delmer, D.P. & Lamport, D.T.A. (1990) Adaptation and growth of tomato cells on the herbicide 2,6-dichlorobenzonitrile leads to production of unique cell walls virtually lacking a cellulose-xyloglucan network. *Plant Physiology*, **94**, 980–987.
329. His, I., Driouich, A., Nicol, F., Jauneau, A. & Höfte, H. (2001) Altered pectin composition in primary cell walls of korrigan, a dwarf mutant of *arabidopsis* deficient in a membrane-bound endo-1,4-beta-glucanase. *Planta*, **212**, 348–358.
330. Piro, G. & Dalessandro, G. (1998) Cell-wall biosynthesis in differentiating cells of pine root tips. *Phytochemistry*, **47**, 1201–1206.

331. Rizk, S.E., Abdel-Massih, R.M., Baydoun, E.A.H. & Brett, C.T. (2000) Protein- and pH-dependent binding of nascent pectin and glucuronoarabinoxylan to xyloglucan in pea. *Planta*, **211**, 423–429.
332. Tavares, R., Aubourg, S., Lechamy, A. & Kreis, M. (2000) Organization and structural evolution of four multigene families in *Arabidopsis thaliana*: AtLCAD, AtLGT, AtMYST and AtHD-GL2. *Plant Molecular Biology*, **42**, 703–717.
333. Bruyant-Vannier, M.-P., Gaudinet-Schaumann, A., Bourlard, T. & Morvan, C. (1996) Solubilization and partial characterization of pectin methyltransferase from flax cells. *Plant Physiology and Biochemistry*, **34**, 489–499.
334. Bourlard, T., Pellerin, P. & Morvan, C. (1997) Rhamnogalacturonans I and II are pectic substrates for flax-cell methyltransferases. *Plant Physiology and Biochemistry*, **35**, 623–629.
335. Goubet, F. & Mohnen, D. (1999) Solubilization and partial characterization of homogalacturonan-methyltransferase from microsomal membranes of suspension-cultured tobacco cells. *Plant Physiology*, **121**, 281–290.
336. Baydoun, E.A.H., Rizk, S.E. & Brett, C.T. (1999) Localisation of methyltransferases involved in glucuronoxylan and pectin methylation in the Golgi apparatus in etiolated pea epicotyls. *Journal of Plant Physiology*, **155**, 240–244.
337. Bourlard, T., Vannier, M.-P., Schaumann, A., Bruyant, P. & Morvan, C. (2001) Purification of several pectin methyltransferases from cell suspension cultures of flax (*Linum usitatissimum* L.). *C.R. Acad. Sci. Paris, Sciences de la vie/Life Sciences*, **324**, 335–343.
338. Lamblin, F., Saladin, G., Dehorter, B., Cronier, D., Grenier, E., Lacoux, J., Bruyant, P., Lainé, E., Chabbert, B., Girault, F., Monties, B., Morvan, C. David, H. & David, A. (2001) Overexpression of a heterologous *sam* gene encoding S-adenosylmethionine synthetase in flax (*Linum usitatissimum*) cells: Consequences on methylation of lignin precursors and pectins. *Plant Physiology*, **112**, 223–232.
339. Pauly, M. & Scheller, H.V. (2000) O-acetylation of plant cell wall polysaccharides: Identification and partial characterization of rhamnogalacturonan O-acetyl-transferase from potato suspension cultured cells. *Planta*, **210**, 659–667.
340. Renard, C.M.G.C., Weightman, R.M. & Thibault, J.-F. (1997) The xylose-rich pectins from pea hulls. *International Journal of Biological Macromolecules*, **21**, 155–162.
341. Pan, Y.-T. & Kindel, P.K. (1977) Characterization of particulate D-apiosyl- and D-xylosyltransferase from *Lemna minor*. *Archives of Biochemistry and Biophysics*, **183**, 131–138.
342. Mascaro, L.J., Jr. & Kindel, P.K. (1977) Characterization of [¹⁴C]apiogalacturonans synthesized in a cell-free system from *Lemna minor*. *Archives of Biochemistry and Biophysics*, **183**, 139–148.
343. Longland, J.M., Fry, S.C. & Trewavas, A.J. (1989) Developmental control of Apiogalacturonan biosynthesis and UDP-Apiose production in a Duckweed. *Plant Physiology*, **90**, 972–976.
344. Spellman, M.W., McNeil, M., Darvill, A.G., Albersheim, P. & Henrick, K. (1983) Isolation and characterization of 3-C-carboxy-5-deoxy-L-xylose, a naturally occurring, branched-chain, acidic monosaccharide. *Carbohydrate Research*, **122**, 115–129.
345. York, W.S., Darvill, A.G., McNeil, M. & Albersheim, P. (1985) 3-Deoxy-D-manno-2-octulosonic acid (KDO) is a component of rhamnogalacturonan II, a pectic polysaccharide in the primary cell walls of plants. *Carbohydrate Research*, **138**, 109–126.
346. Stevenson, T.T., Darvill, A.G. & Albersheim, P. (1988) 3-Deoxy-D-lyxo-2-heptulosaric acid, a component of the plant cell-wall polysaccharide rhamnogalacturonan-II. *Carbohydrate Research*, **179**, 269–288.
347. Spellman, M.W., McNeil, M., Darvill, A.G., Albersheim, P. & Dell, A. (1983) Characterization of a structurally complex heptasaccharide isolated from the pectic polysaccharide rhamnogalacturonan II. *Carbohydrate Research*, **122**, 131–153.
348. Melton, L.D., McNeil, M., Darvill, A.G., Albersheim, P. & Dell, A. (1986) Structural characterization of oligosaccharides isolated from the pectic polysaccharide rhamnogalacturonan II. *Carbohydrate Research*, **146**, 279–305.

349. Stevenson, T.T., Darvill, A.G. & Albersheim, P. (1988) Structural features of the plant cell-wall polysaccharide rhamnogalacturonan-II. *Carbohydrate Research*, **182**, 207–226.
350. Puvanesarajah, V., Darvill, A.G. & Albersheim, P. (1991) Structural characterization of two oligosaccharide fragments formed by the selective cleavage of rhamnogalacturonan II: Evidence for the anomeric configuration and attachment sites of apiose and 3-deoxy-2-heptulosaric acid. *Carbohydrate Research*, **218**, 211–222.
351. Whitcombe, A.J., O'Neill, M.A., Steffan, W., Albersheim, P. & Darvill, A.G. (1995) Structural characterization of the pectic polysaccharide, Rhamnogalacturonan-II. *Carbohydrate Research*, **271**, 15–29.
352. O'Neill, M.A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A.G. & Albersheim, P. (1996) Rhamnogalacturonan-II, a pectic polysaccharide in the walls of growing plant cell, forms a dimer that is covalently cross-linked by a borate ester – in vitro conditions for the formation and hydrolysis of the dimer. *Journal of Biological Chemistry*, **271**, 22923–22930.
353. Matoh, T., Kawaguchi, S. & Kobayashi, M. (1996) Ubiquity of a borate-rhamnogalacturonan II complex in the cell walls of higher plants. *Plant and Cell Physiology*, **37**, 636–640.
354. Kobayashi, M., Matoh, T. & Azuma, J.-I. (1996) Two chains of rhamnogalacturonan II are cross-linked by borate-diol ester bonds in higher plant cell walls. *Plant Physiology*, **110**, 1017–1020.
355. Ishii, T. & Matsunaga, W. (1996) Isolation and characterization of a boron-rhamnogalacturonan-II complex from cell-walls of sugar-beet pulp. *Carbohydrate Research*, **284**, 1–9.
356. Kaneko, S., Ishii, T. & Matsunaga, T. (1997) A boron-rhamnogalacturonan-II complex from bamboo shoot cell walls. *Phytochemistry*, **44**, 243–248.
357. Takasaki, M., Kawaguchi, S., Kobayashi, M., Takabe, K. & Matoh, T. (1997) Immunocytochemistry of the borate-rhamnogalacturonan II complex in cell walls of radish roots. In: *Boron in Soils and Plants* (eds. R.W. Bell & B. Rerkasem), pp. 243–249. Kluwer Academic Publishers, Dordrecht, the Netherlands.
358. Ahn, J.-W., Verma, R., Kim, M., Lee, J.-Y., Kim, Y.-K., Bang, J.-W., Reiter, W.-D. & Pai, H.-S. (2006) Depletion of UDP-D-apiose/UDP-D-xylose synthases results in rhamnogalacturonan-II deficiency, cell wall thickening, and cell death in higher plants. *Journal of Biological Chemistry*, **281**, 13708–13716.
359. Reithmeier, R.A.F. & Deber, C.M. (1992) Intrinsic membrane protein structure: Principles and prediction. In: *The Structure of Biological Membranes* (ed. P. Yeagle), pp. 337–393. CRC Press, Boca Raton, FL.
360. Darvill, A., McNeil, M. & Albersheim, P. (1978) Structure of plant cell walls: VIII. A new pectic polysaccharide. *Plant Physiology*, **62**, 418–422.
361. McNeil, M., Darvill, A.G. & Albersheim, P. (1980) Structure of plant cell walls. X. Rhamnogalacturonan I. A structurally complex pectic polysaccharide in the walls of suspension-cultured sycamore cells. *Plant Physiology*, **66**, 1128–1134.
362. An, J., Zhang, L., O'Neill, M.A., Albersheim, P. & Darvill, A.G. (1994) Isolation and structural characterization of endo-rhamnogalacturonase-generated fragments of the backbone of rhamnogalacturonan I. *Carbohydrate Research*, **264**, 83–96.
363. Thomas, J.R., Darvill, A.G. & Albersheim, P. (1989) Rhamnogalacturonan I, a pectic polysaccharide that is a component of monocot cell-walls. *Carbohydrate Research*, **185**, 279–305.
364. Bhattacharjee, S.S. & Timell, T.E. (1965) A study of the pectin present in the bark of amabilis fir (*Abies amabilis*). *Canadian Journal of Chemistry*, **43**, 758–765.
365. Clarke, A.E., Anderson, R.L. & Stone, B.A. (1979) Form and function of arabinogalactans and arabinogalactan-proteins. *Phytochemistry*, **18**, 521–540.
366. Fincher, G.B., Stone, B.A. & Clarke, A.E. (1983) Arabinogalactan-proteins: Structure, biosynthesis, and function. *Annual Review of Plant Physiology*, **34**, 47–70.
367. Bacic, A., Du, H., Stone, B.A. & Clarke, A.E. (1996) Arabinogalactan proteins: A family of cell-surface and extracellular matrix plant proteoglycans. *Essays in Biochemistry*, **31**, 91–101.

368. Estévez, J.M., Kieliszewski, M.J., Khitrov, N. & Somerville, C. (2006) Characterization of synthetic hydroxyproline-rich proteoglycans with arabinogalactan protein and extensin motifs in *Arabidopsis*. *Plant Physiology*, **142**, 458–470.
369. Serpe, M.D. & Nothnagel, E.A. (1995) Fractionation and structural characterization of arabinogalactan-proteins from the cell wall of rose cells. *Plant Physiology*, **109**, 1007–1016.
370. Gane, A.M., Craik, D., Munro, S.L.A., Howlett, G.J., Clarke, A.E. & Bacic, A. (1995) Structural analysis of the carbohydrate moiety of arabinogalactan-proteins from stigmas and styles of *Nicotiana glauca*. *Carbohydrate Research*, **277**, 67–85.
371. Freshour, G., Clay, R.P., Fuller, M.S., Albersheim, P., Darvill, A.G. & Hahn, M.G. (1996) Developmental and tissue-specific structural alterations of the cell-wall polysaccharides of *Arabidopsis thaliana* Roots. *Plant Physiology*, **110**, 1413–1429.
372. Steffan, W., Kovac, P., Albersheim, P., Darvill, A.G. & Hahn, M.G. (1995) Characterization of a monoclonal antibody that recognizes an arabinosylated (1-6)- β -D-galactan epitope in plant complex carbohydrates. *Carbohydrate Research*, **275**, 295–307.
373. Fry, S.C. (1986) Cross-linking of matrix polymers in the growing cell walls of angiosperms. *Annual Review of Plant Physiology*, **37**, 165–186.
374. Fry, S.C. (1983) Feruloylated pectins from the primary cell wall: their structure and possible functions. *Planta*, **157**, 111–123.
375. McNab, J.M., Vilemez, C.L. & Albersheim, P. (1968) Biosynthesis of galactan by a particulate preparation from *Phaseolus aureus* seedlings. *Biochemical Journal*, **106**, 355–360.
376. Panayotatos, N. & Vilemez, C.L. (1973) The formation of a β -(1 \rightarrow 4)-D-galactan chain catalysed by a *Phaseolus aureus* enzyme. *Biochemical Journal*, **133**, 263–271.
377. Brickell, L.S. & Reid, J.S.G. (1996) Biosynthesis *in vitro* of pectic (1-4)- β -D-galactan. In: *Pectins and Pectinases* (eds. J. Visser & A.G.J. Voragen), pp. 127–134. Elsevier Science B.V., Amsterdam.
378. Goubet, F. & Morvan, C. (1994) Synthesis of cell wall galactans from flax (*Linum usitatissimum* L.) suspension-cultured cells. *Plant and Cell Physiology*, **35**, 719–727.
379. Goubet, F. & Morvan, C. (1993) Evidence for several galactan synthases in flax (*Linum usitatissimum* L.) suspension-cultured cells. *Plant and Cell Physiology*, **34**, 1297–1303.
380. Goubet, F. (1994) *Etude de la biosynthèse de polysaccharides pariétaux des fibres celluloseuses au cours du développement du lin*. PhD thesis, Université de Rouen.
381. Peugnet, I., Goubet, F., Bruyant-Vannier, M.-P., Thoirion, B., Morvan, C., Schols, H.A. & Voragen, A.G.J. (2001) Solubilization of rhamnogalacturonan I galactosyltransferases from membranes of a flax cell suspension. *Planta*, **213**, 435–445.
382. Geshi, N., Jorgensen, B., Scheller, H.V. & Ulvskov, P. (2000) *In vitro* biosynthesis of 1,4- β -galactan attached to rhamnogalacturonan I. *Planta*, **210**, 622–629.
383. Azadi, P., O'Neill, M.A., Bergmann, C., Darvill, A.G. & Albersheim, P. (1995) The backbone of the pectic polysaccharide rhamnogalacturonan I is cleaved by an *endohydrolase* and an *endolyase*. *Glycobiology*, **5**, 783–789.
384. Geshi, N., Pauly, M. & Ulvskov, P. (2002) Solubilization of galactosyltransferase that synthesizes 1,4-beta-galactan side chains in pectic rhamnogalacturonan I. *Physiologica Plantarum*, **114**, 540–548.
385. Geshi, N., Jorgensen, B. & Ulvskov, P. (2004) Subcellular localization and topology of β (1–4) galactosyltransferase that elongates β (1–4) galactan side chains in rhamnogalacturonan I in potato. *Journal of Planta*, **218**, 862–868.
386. Ishii, T., Ohnishi-Kameyama, M. & Ono, H. (2004) Identification of elongating β -1,4-galactosyltransferase activity in mung bean (*Vigna radiata*) hypocotyls using 2-aminobenzaminated 1,4-linked β -D-galactooligosaccharides as acceptor substrates. *Journal of Planta*, **219**, 310–318.

387. Ishii, T., Ono, H. & Maeda, I. (2005) Assignment of the H and C NMR spectra of 2-aminobenzamide-labeled galacto- and arabinooligosaccharides. *Journal of Wood Science*, **51**, 295–302.
388. Odzuck, W. & Kauss, H. (1972) Biosynthesis of pure araban and xylan. *Phytochemistry*, **11**, 2489–2494.
389. Bolwell, G.P. & Northcote, D.H. (1981) Control of hemicellulose and pectin synthesis during differentiation of vascular tissue in bean (*Phaseolus vulgaris*) callus and in bean hypocotyl. *Planta*, **152**, 225–233.
390. Bolwell, G.P. & Northcote, D.H. (1983) Arabinan synthase and xylan synthase activities of *Phaseolus vulgaris*. Subcellular localization and possible mechanism of action. *Biochemical Journal*, **210**, 497–507.
391. Pauly, M., Porchia, A., Olsen, C.E., Nunan, K.J. & Scheller, H.V. (2000) Enzymatic synthesis and purification of UDP- β -L-arabinopyranose, a substrate for the biosynthesis of a plant polysaccharides. *Analytical Biochemistry*, **278**, 69–73.
392. Zhang, Q. & Liu, H. (2001) Chemical synthesis of UDP- β -L-arabinofuranose and its turnover to UDP- β -L-arabinopyranose by UDP-galactopyranose mutase. *Bioorganic and Medicinal Chemistry Letters*, **11**, 145–149.
393. Konishi, T., Takumi, T., Miyazuki, Y., Ohnishi-Kameyama, M., Hayashi, T., O'Neill, M.A. & Ishii, T. (2007) A plant mutase that interconverts UDP-arabinofuranose and UDP-arabinopyranose. *Glycobiology*, **17**, 345–354.
394. Dhugga, K.S., Ulvskov, P., Gallagher, S.R. & Ray, P.M. (1991) Plant polypeptides reversibly glycosylated by UDP-glucose. Possible components of Golgi β -glucan synthase in pea cells. *Journal of Biological Chemistry*, **266**, 21977–21984.
395. Langeveld, S.M., Vennik, M., Kottenhagen, M., van Wijk, R., Buijk, A., Kijne, J.W. & De Pater, S. (2002) Glucosylation activity and complex formation of two classes of reversibly glycosylated polypeptides. *Plant Physiology*, **129**, 278–289.
396. Drakakaki, G., Zabolina, O., Delgado, I., Roberts, S., Keegstra, K. & Raikel, N. (2006) Arabidopsis reversibly glycosylated polypeptides 1 and 2 are essential for pollen development. *Plant Physiology*, **142**, 1480–1492.
397. Nunan, K.J. & Scheller, H.V. (2003) Solubilization of an arabinan arabinosyltransferase activity from mung bean hypocotyls. *Plant Physiology*, **132**, 331–342.
398. Mouille, G., Witucka-Wall, H., Bruyant, M.-P., Loudet, O., Pelletier, S., Rihouey, C., Lerouxel, O., Lerouge, P., Höfte, H. & Pauly, M. (2006) Quantitative trait loci analysis of primary cell wall composition in Arabidopsis. *Plant Physiology*, **141**, 1035–1044.
399. Egelund, J., Obel, N., Ulvskov, P., Geshi, N., Pauly, M., Bacic, A. & Petersen, B.L. (2007) Molecular characterization of two *Arabidopsis thaliana* glycosyltransferase mutants, *rra1* and *rra2*, which have a reduced residual arabinose content in a polymer tightly associated with the cellulosic wall residue. *Plant Molecular Biology*, **64**, 439–451.
400. Puhlmann, J., Bucheli, E., Swain, M.J., Dunning, N., Albersheim, P., Darvill, A.G. & Hahn, M.G. (1994) Generation of monoclonal antibodies against plant cell-wall polysaccharides I. Characterization of a monoclonal antibody to a terminal α -(1 \rightarrow 2)-linked fucosyl-containing epitope. *Plant Physiology*, **104**, 699–710.
401. Camirand, A., Brummell, D. & MacLachlan, G. (1987) Fucosylation of xyloglucan: Localization of the transferase in dictyosomes of pea stem cells. *Plant Physiology*, **84**, 753–756.
402. Brummell, D.A., Camirand, A. & MacLachlan, G. (1990) Differential distribution of xyloglucan glycosyl transferases in pea Golgi dictyosomes and secretory vesicles. *Journal of Cell Science*, **96**, 705–710.
403. Feingold, D.S. (1982) Aldo (and keto) hexoses and uronic acids. In: *Plant Carbohydrates I* (eds. F.A. Loewus & W. Tanner), pp. 3–76. Springer-Verlag, Berlin, Heidelberg.

404. Seifert, G.J. (2004) Nucleotide sugar interconversions and cell wall biosynthesis: How to bring the inside to the outside. *Journal of Current Opinion in Plant Biology*, **7**, 277–284.
405. Kanjilal-Kolar, S., Basu, S.S., Kanipes, M.I., Guan, Z., Garrett, T.A. & Raetz, C.R. (2006) Expression cloning of three *Rhizobium leguminosarum* lipopolysaccharide core galacturonosyltransferases. *Journal of Biological Chemistry*, **281**, 12865–12878.
406. Kanjilal-Kolar, S. & Raetz, C.R.H. (2006) Dodecaprenyl phosphate-galacturonic acid as a donor substrate for lipopolysaccharide core glycosylation in *Rhizobium leguminosarum*. *Journal of Biological Chemistry*, **281**, 12879–12887.
407. Norambuena, L., Nilo, R., Handford, M., Reyes, F., Marchant, L., Meisel, L. & Orellana, A. (2005) AtUTr2 is an Arabidopsis thaliana nucleotide sugar transporter located in the Golgi apparatus capable of transporting UDP-galactose. *Planta*, **222**, 521–529.
408. Harden, A. & Young, W.J. (1906) The alcoholic ferment of yeast juice. *Proceedings of the Royal Society Series B*, **77**, 405–420.
409. Cori, C.F., Schmidt, G. & Cori, G.T. (1939) The synthesis of a polysaccharide from glucose-1-phosphate in muscle extract. *Science*, **89**, 464–465.
410. Caputto, R., Leloir, L.F., Cardini, C.E. & Paladini, A.C. (1950) Isolation of the coenzyme of the galactose phosphate-glucose phosphate transformation. *Journal of Biological Chemistry*, **184**, 333–350.
411. Neufeld, E.F., Feingold, D.S. & Hassid, W.Z. (1959) Enzymic phosphorylation of D-glucuronic acid by extracts from seedlings of *Phaseolus aureus*. *Archives of Biochemistry and Biophysics*, **83**, 96–100.
412. Neufeld, E.F., Feingold, D.S. & Hassid, W.Z. (1960) Phosphorylation of D-galactose and L-arabinose by extracts from *Phaseolus aureus* seedlings. *Journal of Biological Chemistry*, **235**, 906–909.
413. Kotake, T., Yamaguchi, D., Ohzono, H., Hojo, S., Kaneko, S., Ishida, H. & Tsumuraya, Y. (2004) UDP-sugar pyrophosphorylase with broad substrate specificity toward various monosaccharide 1-phosphates from pea sprouts. *Journal of Biological Chemistry*, **279**, 45728–45736.
414. Bar-Peled, L. (2005) *A Novel Pathway for Polysaccharide Precursor Synthesis*. University of Georgia, Center for Undergraduate Research Opportunities (CURO symposium).
415. Schnurr, J.A., Storey, K.K., Jung, H.-J.G., Somers, D.A. & Gronwald, J.W. (2006) UDP-sugar pyrophosphorylase is essential for pollen development in Arabidopsis. *Planta*, **224**, 520–532.
416. Klepek, Y.-S., Geiger, D., Stadler, R., Klebl, F., Landouar-Arsivaud, L., Lemoine, R., Hedrich, R. & Sauer, N. (2005) Arabidopsis POLYOL TRANSPORTER5, a new member of the monosaccharide transporter-like superfamily, mediates H⁺-symport of numerous substrates, including myo-inositol, glycerol, and ribose. *Plant Cell*, **17**, 204–218.
417. Chen, R., Zhao, X., Shao, Z., Wei, Z., Wang, Y., Zhu, L., Zhao, J., Sun, M., He, R. & He, G. (2007) Rice UDP-Glucose pyrophosphorylase1 is essential for pollen callose deposition and its cosuppression results in a new type of thermosensitive genic male sterility. *Plant Cell*, **19**, 847–861.
418. Baroja-Fernández, E., Muñoz, F.J., Saikusa, T., Rodríguez-López, M., Akazawa, T. & Pozueta-Romero, J. (2003) Sucrose synthase catalyzes the de novo production of ADPglucose linked to starch biosynthesis in heterotrophic tissues of plants. *Plant and Cell Physiology*, **44**, 500–509.
419. Delmer, D.P. & Albersheim, P. (1970) The biosynthesis of sucrose and nucleoside diphosphate glucoses in *phaseolus aureus*. *Plant Physiology*, **45**, 782–786.
420. Barratt, D.H.P., Barber, L., Kruger, N.J., Smith, A.M., Wang, T.L. & Martin, C. (2001) Multiple, distinct isoforms of sucrose synthase in pea. *Plant Physiology*, **127**, 655–664.
421. Baud, S., Vaultier, M.-N. & Rochat, C. (2004) Structure and expression profile of the sucrose synthase multigene family in Arabidopsis. *Journal of Experimental Botany*, **55**, 397–409.

422. Winter, H. & Huber, S.C. (2000) Regulation of sucrose metabolism in higher plants: Localization and regulation of activity of key enzymes. *Critical Reviews in Biochemistry and Molecular Biology*, **35**, 253–289.
423. Etxeberria, E. & Gonzalez, P. (2003) Evidence for a tonoplast-associated form of sucrose synthase and its potential involvement in sucrose mobilization from the vacuole. *Journal of Experimental Botany*, **54**, 1407–1414.
424. Subbaiah, C.C., Palaniappan, A., Duncan, K., Rhoads, D.M., Huber, S.C. & Sachs, M.M. (2006) Mitochondrial localization and putative signaling function of sucrose synthase in maize. *Journal of Biological Chemistry*, **281**, 15625–15635.
425. Hardin, S.C., Winter, H. & Huber, S.C. (2004) Phosphorylation of the amino terminus of maize sucrose synthase in relation to membrane association and enzyme activity. *Plant Physiology*, **134**, 1427–1438.
426. Field, R.A. & Naismith, J.H. (2003) Structural and mechanistic basis of bacterial sugar nucleotide-modifying enzymes. *Biochemistry*, **42**, 7637–7647.
427. Thoden, J.B., Wohlers, T.M., Fridovich-Keil, J.L. & Holden, H.M. (2000) Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase. *Biochemistry*, **39**, 5691–5701.
428. Dormann, P. & Benning, C. (1998) The role of UDP-glucose epimerase in carbohydrate metabolism of *Arabidopsis*. *Plant Journal*, **13**, 641–652.
429. Zhang, Q., Hrmova, M., Shirley, N.J., Lahnstein, J. & Fincher, G.B. (2006) Gene expression patterns and catalytic properties of UDP-D-glucose 4-epimerases from barley (*Hordeum vulgare* L.). *Biochemical Journal*, **394**, 115–124.
430. Barber, C., Rösti, J., Rawat, A., Findlay, K., Roberts, K. & Seifert, G.J. (2006) Distinct properties of the five UDP-D-glucose/UDP-D-galactose 4-epimerase isoforms of *Arabidopsis thaliana*. *Journal of Biological Chemistry*, **281**, 17276–17285.
431. Harper, A.D. & Bar-Peled, M. (2002) Biosynthesis of UDP-Xylose. Cloning and characterization of a novel *Arabidopsis* gene family, *UXS*, encoding soluble and putative membrane-bound UDP-glucuronic acid decarboxylase isoforms. *Plant Physiology*, **130**, 2188–2198.
432. Baskin, T.I., Bertzner, A.S., Hoggart, R., Cork, A. & Williamson, R.E. (1992) Root morphology mutants in *Arabidopsis thaliana*. *Australian Journal of Plant Physiology*, **19**, 427–437.
433. Ding, L. & Zhu, J.K. (1997) A role of arabinogalactan-proteins in root epidermal cell expansion. *Planta*, **203**, 289–294.
434. Schiefelbein, J.W. & Somerville, C. (1990) Genetic control of root hair development in *Arabidopsis thaliana*. *Plant Cell*, **2**, 235–243.
435. Seifert, G.J., Barber, C., Wells, B., Dolan, L. & Roberts, K. (2002) Galactose biosynthesis in *Arabidopsis*: Genetic evidence for substrate channeling from UDP-D-galactose into cell wall polymers. *Current Biology*, **12**, 1840–1845.
436. Espada, J. (1962) Enzymatic synthesis of adenosine diphosphate glucose from glucose 1-phosphate and adenosine triphosphate. *Journal of Biological Chemistry*, **237**, 3577–3581.
437. Okita, T.W., Greenberg, E., Kuhn, D.N. & Preiss, J. (1979) Subcellular localization of the starch degradative and biosynthetic enzymes of spinach leaves. *Plant Physiology*, **64**, 187–192.
438. Muñoz, F.J., Baroja-Fernández, E., Morán-Zorzano, M.T., Viale, A.M., Etxeberria, E., Alonso-Casajús, N. & Pozueta-Romero, J. (2005) Sucrose synthase controls both intracellular ADP glucose levels and transitory starch biosynthesis in source leaves. *Plant and Cell Physiology*, **46**, 1366–1376.
439. Munoz, F.J., Moran-Zorzano, M.T., Alonso-Casajús, N., Baroja-Fernández, E., Etxeberria, E. & Pozueta-Romero, J. (2006) New enzymes, new pathways and an alternative view on starch biosynthesis in both photosynthetic and heterotrophic tissues of plants. *Biocatalysis and Bio-transformation*, **24**, 63–76.

440. Crevillén, P., Ballicora, M.A., Mérida, A., Preiss, J. & Romero, J.M. (2003) The different large subunit isoforms of *Arabidopsis thaliana* ADP-glucose pyrophosphorylase confer distinct kinetic and regulatory properties to the heterotetrameric enzyme. *Journal of Biological Chemistry*, **278**, 28508–28515.
441. Crevillén, P., Ventriglia, T., Pinto, F., Orea, A., Mérida, Á. & Romero, J.M. (2005) Differential pattern of expression and sugar regulation of *Arabidopsis thaliana* ADP-glucose pyrophosphorylase-encoding genes. *Journal of Biological Chemistry*, **280**, 8143–8149.
442. Patron, N.J., Greber, B., Fahy, B.F., Laurie, D.A., Parker, M.L. & Denyer, K. (2004) The *lys5* mutations of barley reveal the nature and importance of plastidial ADP-Glc transporters for starch synthesis in cereal endosperm. *Plant Physiology*, **135**, 2088–2097.
443. McCoy, J.G., Arabshahi, A., Bitto, E., Bingman, C.A., Ruzicka, F.J., Frey, P.A. & Phillips, G.N., Jr. (2006) Structure and mechanism of an ADP-glucose phosphorylase from *Arabidopsis thaliana*. *Biochemistry*, **45**, 3154–3162.
444. Kaplan, C.P., Tugal, H.B. & Baker, A. (1997) Isolation of a cDNA encoding an Arabidopsis galactokinase by functional expression in yeast. *Plant Molecular Biology*, **34**, 497–506.
445. Sherson, S., Gy, I., Medd, J., Schmidt, R., Dean, C., Kreis, M., Lecharny, A. & Cobbett, C. (1999) The arabinose kinase, ARA1, gene of Arabidopsis is a novel member of the galactose kinase gene family. *Plant Molecular Biology*, **39**, 1003–1012.
446. Thoden, J.B. & Holden, H.M. (2005) The molecular architecture of human N-acetylgalactosamine kinase. *Journal of Biological Chemistry*, **280**, 32784–32791.
447. Timson, D.J. & Reece, R.J. (2003) Sugar recognition by human galactokinase. *BMC Biochemistry*, **4**, 16.
448. Bar-Peled, M., Lewinsohn, E., Fluhr, R. & Gressel, J. (1991) UDP-rhamnose:flavanone-7-O-glucoside-2'-O-rhamnosyltransferase. Purification and characterization of an enzyme catalyzing the production of bitter compounds in citrus. *Journal of Biological Chemistry*, **266**, 20953–20959.
449. Western, T.L., Young, D.S., Dean, G.H., Tan, W.L., Samuels, A.L. & Haughn, G.W. (2004) MUCILAGE-MODIFIED4 encodes a putative pectin biosynthetic enzyme developmentally regulated by APETALA2, TRANSPARENT TESTA GLABRA1, and GLABRA2 in the Arabidopsis seed coat. *Plant Physiology*, **134**, 296–306.
450. Usadel, B., Schlüter, U., Molhoj, M., Gipmans, M., Verma, R., Kossmann, J., Reiter, W.-D. & Pauly, M. (2004) Identification and characterization of a UDP-D-glucuronate 4-epimerase in *Arabidopsis*. *FEBS Letters*, **569**, 327–331.
451. Oka, T., Nemoto, T. & Jigami, Y. (2007) Functional analysis of *Arabidopsis thaliana* RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. *Journal of Biological Chemistry*, **282**, 5389–5403.
452. Kroh, M. & Loewus, F. (1968) Biosynthesis of pectic substance in germinating pollen: Labeling with myo-inositol-2¹⁴C. *Science*, **160**, 1352–1354.
453. Kanter, U., Usadel, B., Guerineau, F., Li, Y., Pauly, M. & Tenhaken, R. (2005) The inositol oxygenase gene family of *Arabidopsis* is involved in the biosynthesis of nucleotide sugar precursors for cell-wall matrix polysaccharides. *Planta*, **221**, 243–254.
454. Johnson, M.D. & Sussex, I.M. (1995) 1L-myo-inositol 1-phosphate synthase from *Arabidopsis thaliana*. *Plant Physiology*, **107**, 613–619.
455. Parthasarathy, L., Vadnal, R.E., Parthasarathy, R. & Devi, C.S. (1994) Biochemical and molecular properties of lithium-sensitive myo-inositol monophosphatase. *Life Sciences*, **54**, 1127–1142.
456. Gillaspay, G.E., Keddie, J.S., Oda, K. & Gruissem, W. (1995) Plant inositol monophosphatase is a lithium-sensitive enzyme encoded by a multigene family. *Plant Cell*, **7**, 2175–2185.
457. Laing, W.A., Bulley, S., Wright, M., Cooney, J., Jensen, D., Barraclough, D. & MacRae, E. (2004) A highly specific L-galactose-1-phosphate phosphatase on the path to ascorbate biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 16976–16981.

458. Conklin, P.L., Gatzek, S., Wheeler, G.L., Dowdle, J., Raymond, M.J., Rolinski, S., Isupov, M., Littlechild, J.A. & Smirnoff, N. (2006) *Arabidopsis thaliana* VTC4 encodes L-galactose 1-P phosphatase, a plant ascorbic acid biosynthetic enzyme. *Journal of Biological Chemistry*, **281**, 15662–15670.
459. Kessler, G., Neufeld, E.F., Feingold, D.S. & Hassid, W.Z. (1961) Metabolism of D-glucuronic acid and D-galacturonic acid by *Phaseolus aureus* seedlings. *Journal of Biological Chemistry*, **236**, 308–312.
460. Tenhaken, R. & Thulke, O. (1996) Cloning of an enzyme that synthesizes a key nucleotide-sugar precursor of hemicellulose biosynthesis from soybean: UDP-glucose dehydrogenase. *Plant Physiology*, **112**, 1127–1134.
461. Seitz, B., Klos, C., Wurm, M. & Tenhaken, R. (2000) Matrix polysaccharide precursors in *Arabidopsis* cell walls are synthesized by alternate pathways with organ-specific expression patterns. *Plant Journal*, **21**, 537–546.
462. Johansson, H., Sterky, F., Amini, B., Lundeborg, J. & Kleczkowski, L.A. (2002) Molecular cloning and characterization of a cDNA encoding popular UDP-glucose, a key gene of hemicellulose/pectin formation. *Biochimica et Biophysica Acta*, **1576**, 53–58.
463. Bindschedler, L.V., Wheatley, E., Gay, E., Cole, J., Cottage, A. & Bolwell, G.P. (2005) Characterisation and expression of the pathway from UDP-glucose to UDP-xylose in differentiating tobacco tissue. *Plant Molecular Biology*, **57**, 285–301.
464. Kärkönen, A. & Fry, S.C. (2006) Novel characteristics of UDP-glucose dehydrogenase activities in maize: non-involvement of alcohol dehydrogenases in cell wall polysaccharide biosynthesis. *Planta*, **223**, 858–870.
465. Qinghua, H., Dairong, Q., Qinglian, Z., Shunji, H., Yin, L., Linhan, B., Zhirong, Y. & Yi, C. (2005) Cloning and expression studies of the *Dunaliella salina* UDP-glucose dehydrogenase cDNA. *DNA sequence*, **16**, 202–206.
466. Turner, W. & Botha, F.C. (2002) Purification and kinetic properties of UDP-glucose dehydrogenase from sugarcane. *Archives of Biochemistry and Biophysics*, **407**, 209–216.
467. Robertson, D., Smith, C. & Bolwell, G.P. (1996) Inducible UDP-glucose dehydrogenase from French bean (*Phaseolus vulgaris* L.) locates to vascular tissue and has alcohol dehydrogenase activity. *Biochemical Journal*, **313**, 311–317.
468. Neufeld, E.F., Feingold, D.S., Ilves, S.M., Kessler, G. & Hassid, W.Z. (1961) Phosphorylation of D-galacturonic acid by extracts from germinating seeds of *Phaseolus aureus*. *Journal of Biological Chemistry*, **236**, 3102–3105.
469. Ohashi, T., Cramer, N., Ishimizu, T. & Hase, S. (2006) Preparation of UDP-galacturonic acid using UDP-sugar pyrophosphorylase. *Analytical Biochemistry*, **352**, 182–187.
470. Agius, F., Gonzalez-Lamothe, R., Caballero, J.L., Munoz-Blanco, J., Botella, M.A. & Valpuesta, V. (2003) Engineering increased vitamin C levels in plants by overexpression of a D-galacturonic acid reductase. *Nature Biotechnology*, **21**, 177–181.
471. Gu, X. & Bar-Peled, M. (2004) The biosynthesis of UDP-galacturonic acid in plants. Functional cloning and characterization of *Arabidopsis* UDP-D-glucuronic acid 4-epimerase. *Plant Physiology*, **136**, 4256–4264.
472. Molhoj, M., Verma, R. & Reiter, W.-D. (2004) The biosynthesis of D-galacturonate in plants. Functional cloning and characterization of a membrane-anchored UDP-D-glucuronate 4-epimerase from *Arabidopsis*. *Plant Physiology*, **135**, 1221–1230.
473. Pattathil, S., Harper, A.D. & Bar-Peled, M. (2005) Biosynthesis of UDP-xylose: Characterization of membrane-bound At xs2. *Planta*, **221**, 538–548.
474. Suzuki, K., Watanabe, K., Masamura, T. & Kitamura, S. (2004) Characterization of soluble and putative membrane-bound UDP-glucuronic acid decarboxylase (OsUXS) isoforms in rice. *Archives of Biochemistry and Biophysics*, **431**, 169–177.
475. Zhang, Q., Shirley, N., Lahnstein, J. & Fincher, G.B. (2005) Characterization and expression patterns of UDP-D-glucuronate decarboxylase genes in barley. *Plant Physiology*, **138**, 131–141.

476. Gebb, C., Baron, D. & Grisebach, H. (1975) Spectroscopic evidence for the formation of a 4-keto intermediate in the UDP-apiose/UDP-xylose synthase reaction. *European Journal of Biochemistry*, **54**, 493–498.
477. Sandermann, H., Jr., Tissue, G.T. & Grisebach, H. (1968) Biosynthesis of D-apiose. IV. Formation of UDP-apiose from UDP-D-glucuronic acid in cell-free extracts of parsley (*Apium petroselinum* L.) and *Lemna minor*. *Biochimica et Biophysica Acta*, **165**, 550–552.
478. Molhoj, M., Verma, R. & Reiter, W.-D. (2003) The biosynthesis of the branched-chain sugar D-apiose in plants: Functional cloning and characterization of a UDP-D-apiose/UDP-D-xylose synthase from *Arabidopsis*. *Plant Journal*, **35**, 693–703.
479. Dolezal, O. & Cobbett, C.S. (1991) Arabinose kinase-deficient mutant of *Arabidopsis thaliana*. *Plant Physiology*, **96**, 1255–1260.
480. Burget, E.G., Verma, R., Molhoj, M. & Reiter, W.-D. (2003) The biosynthesis of l-arabinose in plants: Molecular cloning and characterization of a Golgi-localized UDP-D-xylose 4-epimerase encoded by the *MUR4* gene of *Arabidopsis*. *Plant Cell*, **15**, 523–531.
481. Feyerabend, D.S., Li, J.P., Lindahl, U. & Rodewald, H.R. (2006) Heparan sulfate C5-epimerase is essential for heparin biosynthesis in mast cells. *Nature Chemical Biology*, **2**, 195–196.
482. Delgado, I.J., Wang, Z., De Rocher, A., Keegstra, K. & Raikhel, N.V. (1998) Cloning and characterization of AtRGP1. A reversibly autoglycosylated arabidopsis protein implicated in cell wall biosynthesis. *Plant Physiology*, **116**, 1339–1349.
483. Dhugga, K.S., Tiwari, S.C. & Ray, P.M. (1997) A reversibly glycosylated polypeptide (RGP1) possibly involved in plant cell wall synthesis: Purification, gene cloning, and trans-Golgi localization. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 7679–7684.
484. Sagi, G., Katz, A., Guenoune-Gelbart, D. & Epel, B.L. (2005) Class 1 reversibly glycosylated polypeptides are plasmodesmal-associated proteins delivered to plasmodesmata via the Golgi apparatus. *Plant Cell*, **17**, 1788–1800.
485. Cabib, E. & Leloir, L.F. (1954) Guanosine diphosphate mannose. *Journal of Biological Chemistry*, **206**, 779–790.
486. Carlson, D.M. & Hansen, R.G. (1962) The isolation and synthesis of guanosine diphosphate glucose. *Journal of Biological Chemistry*, **237**, 1260–1265.
487. Lukowitz, W., Nickle, T.C., Meinke, D.W., Last, R.L., Conklin, P.L. & Somerville, C.R. (2001) *Arabidopsis* *cyt1* mutants are deficient in a mannose-1-phosphate guanylyltransferase and point to a requirement of N-linked glycosylation for cellulose biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **27**, 2262–2267.
488. Conklin, P.L., Norris, S.R., Wheeler, G.L., Williams, E.H., Smirnoff, N. & Last, R.L. (1999) Genetic evidence for the role of GDP-mannose in plant ascorbic acid (vitamin C) biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 4198–4203.
489. Conklin, P.L., Saracco, S.A., Norris, S.R. & Last, R.L. (2000) Identification of ascorbic acid-deficient *Arabidopsis thaliana* mutants. *Genetics*, **154**, 847–856.
490. Bonin, C.P., Freshour, G., Hahn, M.G., Vanzin, G.F. & Reiter, W.-D. (2003) The GMD1 and GMD2 genes of *Arabidopsis* encode isoforms of GDP-D-mannose 4,6-dehydratase with cell type-specific expression patterns. *Plant Physiology*, **132**, 883–892.
491. Bonin, C.P., Potter, I., Vanzin, G.F. & Reiter, W.-D. (1997) The *MUR1* gene of *Arabidopsis thaliana* encodes an isoform of GDP-D-mannose-4,6-dehydratase, catalyzing the first step in the *de novo* synthesis of GDP-L-fucose. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 2085–2090.
492. Reiter, W.-D., Chapple, C.C.S. & Somerville, C.R. (1993) Altered growth and cell walls in a fucose-deficient mutant of *Arabidopsis*. *Science*, **261**, 1032–1035.

493. Goudsmit, E.M. & Neufeld, E.F. (1967) Formation of GDP-L-galactose from GDP-D-mannose. *Biochemical and Biophysical Research Communications*, **26**, 730–735.
494. Wolucka, B.A., Persiau, G., Van Doorselaere, J., Davey, M.W., Demol, H., Vandekerckhove, J., Van Montagu, M., Zabeau, M. & Boerjan, W. (2001) Partial purification and identification of GDP-mannose 3',5''-epimerase of *Arabidopsis thaliana*, a key enzyme of the plant vitamin C pathway. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 14843–14848.
495. Watanabe, K., Suzuki, K. & Kitamura, S. (2006) Characterization of a GDP-D-mannose 3'',5''-epimerase from rice. *Phytochemistry*, **67**, 338–346.
496. Wolucka, B.A. & Van Montagu, M. (2003) GDP-mannose 3',5'-epimerase forms GDP-L-glucose, a putative intermediate for the *de novo* biosynthesis of vitamin C in plants. *Journal of Biological Chemistry*, **278**, 47483–47490.
497. Major, L.L., Wolucka, B.A. & Naismith, J.H. (2005) Structure and function of GDP-mannose-3',5'-epimerase: An enzyme which performs three chemical reactions at the same active site. *Journal of the American Chemical Society*, **127**, 18309–18320.
498. Brabetz, W., Wolter, F.P. & Brade, H. (2000) A cDNA encoding 3-deoxy-D-manno-oct-2-ulosonate-8-phosphate synthase of *Pisum sativum* L. (pea) functionally complements a *kdsA* mutant of the Gram-negative bacterium *Salmonella enterica*. *Planta*, **212**, 136–143.
499. Delmas, F., Petit, J., Joubes, J., Seveno, M., Paccalet, T., Hernould, M., Lerouge, P., Mouras, A. & Chevalier, C. (2003) The gene expression and enzyme activity of plant 3-deoxy-D-manno-2-octulosonic acid- β -phosphate synthase are preferentially associated with cell division in a cell cycle-dependent manner. *Plant Physiology*, **133**, 348–360.
500. Matura, K., Miyagawa, I., Kobayashi, M., Ohta, D. & Matoh, T. (2003) Arabidopsis 3-deoxy-D-manno-oct-2-ulosonate-8-phosphate synthase: cDNA cloning and expression analyses. *Journal of Experimental Botany*, **54**, 1785–1787.
501. Royo, J., Gomez, E. & Hueros, G. (2000) A maize homologue of the bacterial CMP-3-deoxy-D-manno-2-octulosonate (KDO) synthetases. Similar pathways operate in plants and bacteria for the activation of KDO prior to its incorporation into outer cellular envelopes. *Journal of Biological Chemistry*, **275**, 24993–24999.
502. Linster, C.L., Gomez, T.A., Christensen, K.C., Adler, L.N., Young, B.D., Brenner, C. & Clarke, S.G. (2007) Arabidopsis VTC2 encodes a GDP-L-galactose phosphorylase, the last known enzyme in the Smirnoff-Wheeler pathway to ascorbic acid in plants. *Journal of Biological Chemistry*, **282**, 18879–18885.
503. Wells, L., Kreppel, L.K., Comer, F.I., Wadzinski, B.E. & Hart, G.W. (2004) O-GlcNAc transferase is in a functional complex with protein phosphatase 1 catalytic subunits. *Journal of Biological Chemistry*, **279**, 38466–38470.
504. Majumder, A.L., Johnson, M.D. & Henry, S.A. (1997) 1L-myo-inositol-1-phosphate synthase. *Biochimica et Biophysica Acta*, **1348**, 245–256.
505. Pear, J.R., Kawagoe, Y., Schreckengost, W.E., Delmer, D.P. & Stalker, D.M. (1996) Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 12637–12642.
506. Taylor, N.G., Scheible, W.-R., Cutler, S., Somerville, C.R. & Turner, S.R. (1999) The *irregular xylem3* locus of *Arabidopsis* encodes a cellulose synthase required for secondary cell wall synthesis. *Plant Cell*, **11**, 769–779.
507. Ye, Z.-H., York, W.S. & Darvill, A.G. (2006) Important new players in secondary wall synthesis. *Trends in Plant Science*, **11**, 162–164.

Chapter 6

Structures of Plant Cell Wall Celluloses

*Rajai H. Atalla, John W. Brady, James F. Matthews,
Shi-You Ding, and Michael E. Himmel*

6.1 Introduction

In this chapter, we focus on the accessibility of native celluloses in plant cell walls to hydrolytic agents and the manner in which accessibility is modified by dehydration and thermal chemical pretreatments. The primary barrier to enzymatic hydrolysis of celluloses in living plants is their encrustation with lignin. To overcome this barrier, it is necessary to remove the lignin chemically or to fragment the tissue to expose the cellulose; that is, the cell wall must be deconstructed chemically or mechanically. Chemical deconstruction is usually carried out at high temperatures. Mechanical deconstruction is also usually carried out at elevated temperatures to facilitate removal of hemicelluloses, if the primary objective is hydrolysis of the cellulose to glucose. It is therefore important to assess the effects of temperature on the state of aggregation of cellulose. Past studies have shown that temperature elevation almost always results in tighter aggregation of the cellulose chains in the microfibrils in a manner that reduces their accessibility to hydrolytic agents. It is important, therefore, to understand the state of aggregation in the living plant prior to dehydration and to better understand the transformations that arise as a consequence of dehydration in the course of pretreatment of plant biomass.

Microscopic evidence suggests that the microfibrils and nanofibrils of cellulose in higher plants possess a long-period helical twist in their native state. Though the microscopic evidence has revealed the twist in bacterial and algal celluloses, recent theoretical analyses indicate that higher plant celluloses also possess a helical twist that is more pronounced because of its shorter period, and this indeed has been confirmed through atomic force microscope (AFM) imaging. It is important, therefore, to review the evidence regarding the helical twist characteristic of the native state and consider how the highly ordered biological structures, which are species- and tissue-specific, are transformed as a consequence of dehydration into states that are less species- and tissue-specific. We will also consider the driving forces responsible for the dehydration. Finally, we will consider how pretreatment processes might be modified to preserve as much as possible the inherent accessibility of celluloses in their native state.

6.2 Background

The prevailing paradigm regarding the structures of native celluloses during most of the last century held that cellulose is inherently crystalline in the native state and that removal of other cell wall constituents results in exposure of the cellulose in its native state. This is best represented by the models of native celluloses presented by Preston (1) and by Frey-Wyssling (2) in their respective classic treatises. The paradigm has been the basis of all crystallographic models of the structures of native celluloses. This view was based on early observations of the birefringence of cellulose, and after X-ray diffraction was discovered, by observations of X-ray diffraction by cellulosic substances. As new instrumental methods have been developed and as progress has been made in computational modeling of celluloses, it has become clear that questions of structure are more complex.

The evolution of crystallographic models of cellulose has been reviewed elsewhere (3). There had been little consensus regarding the crystallographic structures of the native form well into the early 1980s. The reason for the uncertainty was that different investigators had data sets from different tissues and species, and the different investigators used different constraints on their solutions of the structures. The constraints are necessary because the data sets are not adequate for achieving a definitive solution of the structural problem.

In 1984, Atalla and VanderHart (4, 5) reported that native celluloses are composites of two forms, I_α and I_β , which coexist in all native forms. Two new instrumental methods for that time, Raman spectroscopy and solid state ^{13}C NMR, contributed to the finding. In the initial reports, the two forms were described as “two distinct crystalline forms.” In retrospect, use of the term “crystalline” was unfortunate. Raman spectroscopy also led to the conclusion that the two forms have the same conformation but different hydrogen bonding patterns (6).

The crystallographic studies have been limited historically by the inadequacy of the number of reflections observed for a definitive solution. Diffraction patterns include at most about 300 reflections from celluloses, whereas a definitive solution requires many more reflections. The approximately 300 reflections achieved in the most recent studies of cellulose structures (7, 8) are in contrast to 1257 reflections observed in the study of cellobiose by Chu and Jeffrey (9) and 1724 reflections in the study of methyl β -cellobioside by Ham and Williams (10). In the crystallographic studies of cellulose cited above, the reflections are complemented with *constraints imposed on the solutions of the structural problem*. The key constraints are assumptions regarding the symmetry of the crystal structures that have been controversial since Honjo and Watanabe first reported that such symmetry is not consistent with the electron diffraction patterns (11).

Regarding the organization of celluloses in their native states in higher plants, the central flaw in crystallographic studies is that the constraint represented by the assumption of translational symmetry, implicitly imposed in the mathematical analysis of diffractometric data, is not consistent with the curvature of the microfibrils nor does it accommodate the naturally occurring long-period helical twist. Both the curvature and the twist are distinctive of the morphology of the cell walls and of the tissues and species where they occur. To understand the genomic encoding of species and tissue specificity, it is necessary to have a structural paradigm that can be related to the variability in morphology.

Structures derived from diffractometric measurements require that the forms I_α and I_β belong to different space groups. I_α must have one chain per unit cell with a distinctive conformation, and I_β has two nonequivalent chains per unit cell distinct from the chain in the I_α form. These findings cannot be reconciled with the two forms of native cellulose coexisting in a nanofibril 3–5 nm in diameter. These findings also contradict those by Sugiyama and coworkers in the classic study reporting the lattice images of *Valonia ventricosa* fibrils (12), which shows that the fibril is a single crystal though it is approximately 65% I_α and 35% I_β . Finally, recent observations of the Raman spectra of celluloses I_α and I_β leave little question that the conformations in both forms are nearly identical. These matters have been discussed in detail elsewhere (13); they will be considered in overview here.

From a more conceptual perspective, the fundamental flaw of the crystallographic models is that they do not provide a basis for rationalizing the species and tissue specificity of the blends of the two forms that occur in native tissues. It is clear that the distinctive patterns of different native celluloses are related to the aggregation of the most elementary fibrils at the levels immediately above the aggregation of the individual nanofibrils emerging from individual cellulose synthase complexes (see Chapter 5 for more discussion).

To clarify terminology used in the following discussions, we define nanofibrils as the aggregates of cellulose chains emerging from individual synthase complexes, whereas a microfibril is used to define the next level of aggregation. A microfibril is thus an aggregate of nanofibrils distinctive of a particular tissue in a particular species, and it is usually the first level observable through high magnification microscopy, whether electron microscopy, or during the last decade or so, atomic force microscopy.

The point of departure for our discussion will be the observations by electron microscopy that microfibrils of native celluloses are first and foremost biological structures spatially organized in a periodic helical form. Furthermore, their period varies with the lateral dimensions of the fibrils and is species- and tissue-specific. Recent molecular modeling studies of aggregates of cellulose chains in a hydrated environment have shown that the periodicity is inherent in the nature of cellulose molecules (14).

6.3 Cellulose microfibrils

Here we first review evidence regarding the native state. Perhaps the most informative observations from a quantitative perspective are those of Hanley *et al.* regarding the alga *Micrasterias denticulata* (15) and those of Haigler regarding celluloses formed by the bacterium *Acetobacter xylinum* (16). In both instances, a long-period helical form is observed.

A micrograph of *Micrasterias denticulata* provides an excellent demonstration of the regularity of the periodicity. It is shown in Figure 6.1, where the period of about 1200 nm is evident. Panel A shows the micrographs of the fibril, while panel B provides the authors' rationalization of the appearance of linear segments connected by highly deformed segments within which the 180° turning occurs. The linear segments are about 600 nm each, so that a helical twist of 180° occurs over 600 nm, and two linear segments totaling about 1200 nm corresponding to the full period wherein a complete turn of 360° occurs.

A similar periodicity has been observed by Haigler in her pioneering studies of the structures of bacterial celluloses from *Acetobacter xylinum* and their response to different perturbations of their growth environment (16). The period consistently observed is about 600 nm,

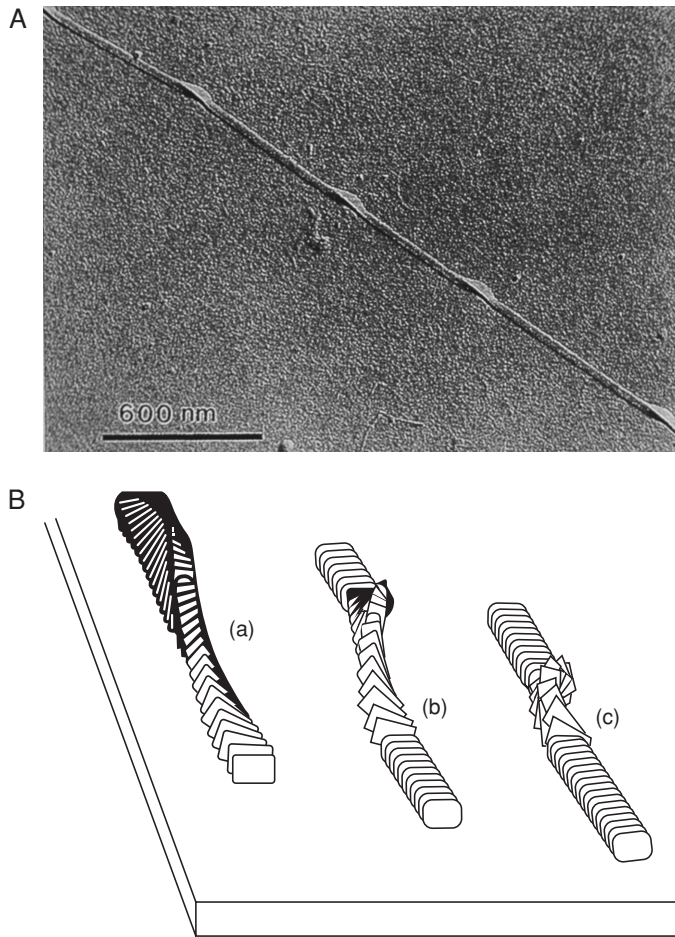


Figure 6.1 A fibril of *Micrasterias denticulata* as observed in panel A, and the rationalization of the appearance of the dehydrated sample (B) proposed by the authors (15). (Courtesy of Professor Derek Gray, Paprican and McGill University, Montreal, Quebec.)

which has also been reported for celluloses formed by *A. xylinum* (17). A key point is that the long period of the helical biological structure appears to depend on lateral dimensions. The lateral dimensions of *Micrasterias denticulata* fibrils are approximately 10 by 20 nm, whereas the lateral dimensions of bacterial cellulose microfibrils are approximately 6–7 nm. Thus, one would expect that the period of the nanofibrils in higher plant cell walls, which have lateral dimensions of the order of 3–5 nm, would be significantly less than 600 nm.

The conclusion regarding the period of higher plant fibrils has been confirmed by the most recent comprehensive molecular modeling studies carried out for hydrated aggregates of cellulose molecules (14). It is also confirmed by atomic force microscopic images of maize parenchyma cell walls shown in Figure 6.2. Here we see that the fibrils appear to vary in width as indicated by the arrows. The variation in width is indicative of an ellipsoidal cross section as

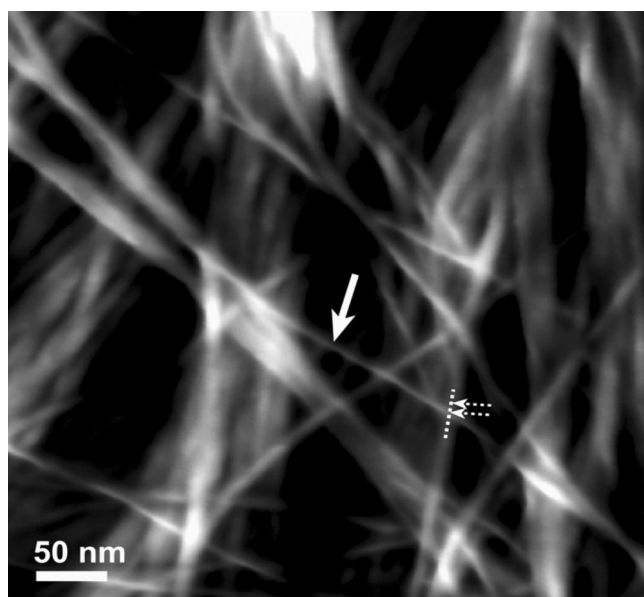


Figure 6.2 An atomic force microscope (AFM) image of the surface of maize parenchyma cell. The large arrow shows a single microfibril at its narrow point. The two smaller arrows show where it is atop another microfibril. Measurement of the elevation provides an approximate dimension. [Adapted from Himmel, M.E., Ding, S.-Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W. & Foust, T.D. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, **315**, 804–807.]

suggested by the ellipsoidal polyhedral model proposed by Ding and Himmel (18). The separation of the narrower sections along an individual fibril appears to be of the order of 200–250 nm. The lateral dimensions of the ellipsoidal polyhedron are approximately 3 by 5 nm.

Before discussing the implications of the molecular modeling studies and patterns of fibril aggregation, it is helpful to clarify the questions regarding symmetry and helical organization with the aid of geometric models of the constrained crystallographic structures and the unconstrained structures manifesting the long-period helical twist. These are illustrated in Figure 6.3, which shows scaled representation of nanofibrils of different sizes. In panel A, they are represented as they would be if they were describable in terms of the symmetry of space groups as implicit in the crystallographically determined structures. They range in cross-sectional size from 2 by 2 nm, which approximates the most elementary fibrils observed, to 20 by 20 nm fibrils representative of algae such as *Valonia* and *Cladophora* as well as the tunicate *Halocynthia*. The length dimension has been scaled to be 300 nm presented in 4 nm intervals to aid in visualization. The geometry of fibrils was defined by the requirement of translational symmetry along three non-coplanar linear axes in Cartesian space. Note that the angles between axes will not be 90° for many crystallographic space groups, although linear translational symmetry in three dimensions is foundational in space group theory (19). Panel B in contrast shows the same fibrils as transformed to reflect a helix with a period of 1200 nm. The figure in panel B represents a 90° turn of the end of each fibril over 300 nm to display the effect of a 360° turn over 1200 nm.

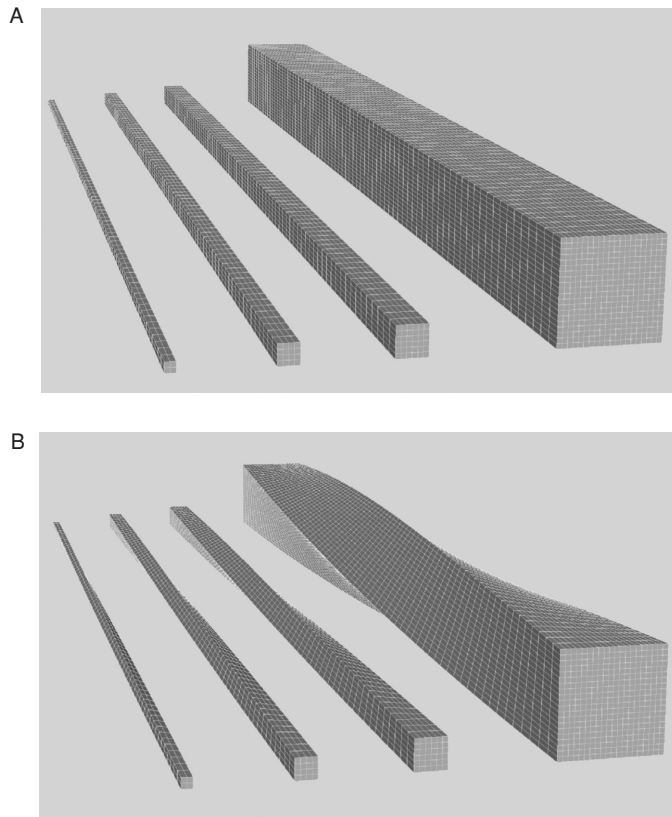


Figure 6.3 Geometric representation of nanofibrils that are 2 by 2, 4 by 4, 6 by 6, and 20 by 20 nm in cross section. Panel A illustrates nanofibrils in translational symmetry in three directions. (Reproduced in color as Plate 5.) Panel B shows a long period of 1200 nm as shown in Figure 6.1. (Reproduced in color as Plate 6.)

It is immediately obvious that assumptions underlying crystallographic analyses are approximations, the implications of which cannot be ignored. For this reason, we need to adopt new terminology to avoid confusion. We suggest no longer using the terms crystal or crystalline, but rather use the term “aggregate” to indicate ordered fibrils. We regard them as highly ordered biological structures that do not meet the classical criteria for crystallinity illustrated in panel A of Figure 6.3. However, since they are periodic at the molecular level as well, we anticipate that they will diffract X-rays in a pattern that approximates that expected from the structures depicted in panel A of Figure 6.3.

The central problem for crystallographic analyses is that the helical twist in fibrils *eliminates the possibility of constructing a reciprocal space*, and such a construction is essential for the interpretation of diffractometric data. Given this observation, one must ask why it has been ignored for approximately 100 years since crystallinity in cellulose was first proposed and diffractometric studies were undertaken. We suggest that a key obstacle has been transformation of native celluloses in the course of isolation. Before reviewing the matter further,

it is useful to begin consideration of the native state in living plants by reviewing the results of the molecular modeling program.

6.3.1 Molecular modeling

Molecular dynamics and molecular mechanics calculations have been used extensively to examine cellulose, often giving unexpected results. Because of inadequate information contained in fiber diffraction patterns noted above, all models for cellulose have been developed to some extent through modeling wherein constraints are imposed on the solution to complement inadequate data sets. The constraints most often used are dimensions of the unit cell together with assumptions regarding symmetry that allow the development of a reciprocal space. These constraints are not unlike boundary conditions necessary for the solution of differential equations in the contexts of mathematical formulations of descriptions of specific physical phenomena.

The molecular dynamics (MD) simulations carried out recently by Matthews and coworkers represent the least constrained analyses (14). The only constraint used was an initial condition corresponding to the published structure of the I_β form. More recently, similar simulations were carried out with the structure of the I_α form as the initial condition and conclusions were essentially the same as those derived from simulation with the structure of the I_β form as the initial condition. During the course of the simulations, a number of structural fluctuations and changes occurred. Over the length of the simulations, average unit cell dimensions shifted away from those reported on the basis of diffractometric measurements; here we use the term unit cell only to provide a basis for comparison with the published structures of the same forms. These dimensions varied with position in the aggregate relative to the surfaces and the chain termini. The results, averaged over these cellobiose units, are summarized in Figure 6.4 and compared with the crystallographic unit cell of the I_β form.

In the simulation, the aggregate was observed to undergo an expansion in which the value of lattice constant a increased from 7.784 to 8.470 Å, while the b value decreased slightly from 8.201 to 8.112 Å. The c value expanded significantly, from 10.380 to 10.512 Å. In addition, the γ angle decreased from 96.5° to almost orthogonal, $\gamma \sim 90^\circ$. The unit cell a -axis (corresponding to the distance between hydrogen bonded sheets) in this simulation differs considerably from that proposed in the crystallographic models. The terminology regarding lattice structures and unit cells is used here primarily to allow comparisons with the structures proposed on the basis of diffractometric measurements.

Another extremely significant change in the structure of the aggregate that occurred during the simulations is that many of the C6 primary alcohol groups underwent rotational transitions away from the conformations reported for the diffraction-based structure. This exocyclic group has three low-energy staggered conformations, which are named TG, GG, and GT, with the first letter in these labels specifying the position of the O6 atom as either *trans* or *gauche* with respect to the O5 atom, and the second letter specifies its relationship to the C4 atom (see Figure 6.5) (20). In both the I_α and I_β diffraction-based structures, all of these exocyclic groups are in the TG conformation. This is also one of the consequences of the constraints of symmetry.

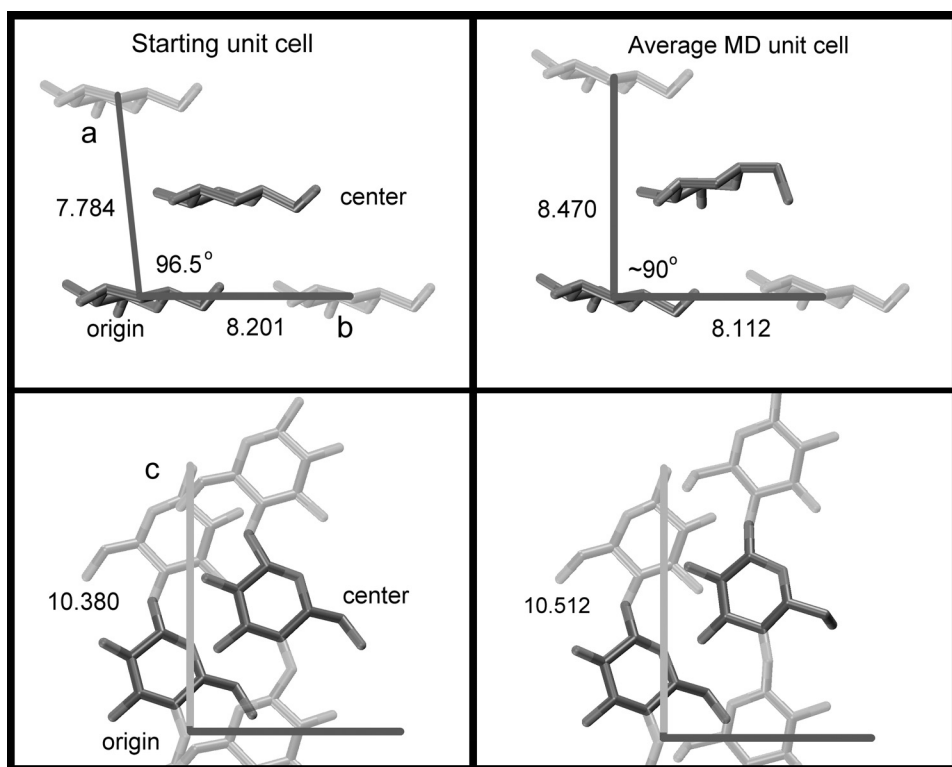


Figure 6.4 Left: the cellulose I_{β} crystal unit cell determined by fiber diffraction; right: the trajectory-averaged unit cell for the simulation of the diagonal crystal. Hydrogen atoms are omitted for clarity, and positions obtained by symmetry operations are transparent. (Reproduced in color as Plate 7.)

In the I_{β} diffraction-based structure, all these exocyclic groups are in the TG conformation. In this conformation, the exocyclic hydroxyl groups can hydrogen-bond along the chain or to adjacent chains in the same layer, but no hydrogen bonds between layers can occur. For those layers of the aggregate made up of the origin chains, there was little change in structure in the MD simulation from that of the diffraction structure, and the hydrogen bonding pattern remained the same. This result is remarkably similar to the reported

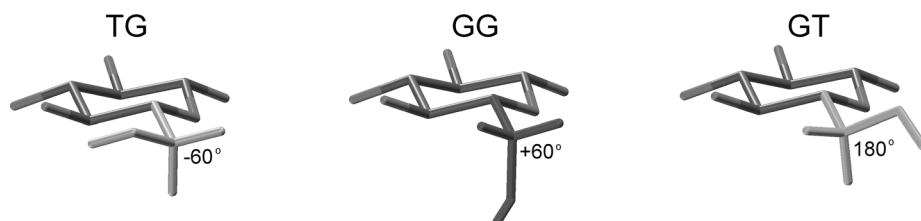


Figure 6.5 Nomenclature of primary alcohol conformation. The dihedral angle measured by C4–C5–C6–O6 is shown. (Reproduced in color as Plate 8.)

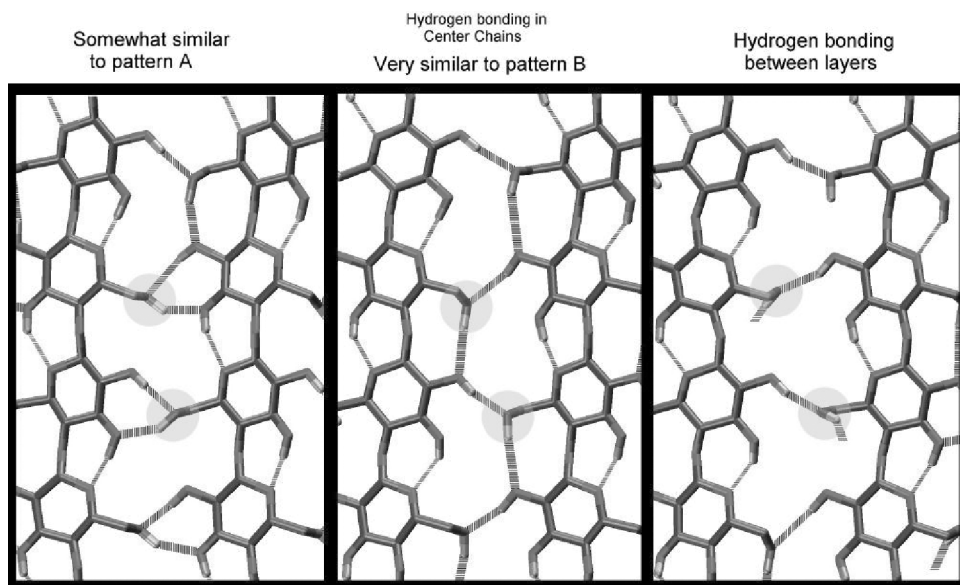


Figure 6.6 Single frames from the center chain layers illustrating three different hydrogen bond patterns. Left: the pattern is similar to the predominant pattern from the crystal structure, but the rotation to GG makes the HO2–O6 hydrogen bond across the glycosidic linkage impossible; center: the hydrogen bond pattern is very similar to the less occupied pattern from the crystal structure; right: hydrogen bonds from HO6 in a center chain to O2 in an origin layer chain, which is not shown for clarity. (Reproduced in color as Plate 9.)

crystallographic hydrogen bond network in origin chains, where the O2 hydroxyl group was refined to just one of the two possible hydrogen bond positions (8). However, in the MD simulations, in every other layer in the interior of the aggregate, made up of the center chains in the diffraction-based structure, this primary alcohol group rotated from the starting TG conformation to the GG position. In this GG conformation, three rapidly interchanging hydrogen bond patterns were possible, as shown in Figure 6.6.

One of these patterns allowed hydrogen bonding between layers, which was not possible when all the hydroxymethyl groups were in the TG conformation. On the surfaces, where the anhydroglucose monomers were in direct contact with water, the hydrogen bonds to the freely diffusing water molecules helped to introduce considerable disorder into these primary alcohol conformations and promoted frequent transitions, but the interior portions of the aggregate developed two distinct patterns of hydroxymethyl conformations between the center and origin layers. Primary alcohol groups in surface chains alternate between facing toward the interior and facing the solvent, and the conformation of these surface groups corresponds to the local environment. Several NMR studies have determined that the conformations of surface cellulose chains are different from the interior, and as in the new simulation, contain both GG and GT rotamers (20–23). The presence of two rotamers is also consistent with the Raman spectra of I_α and I_β to be discussed further in a following section. In the spectra of both I_α and I_β the scissors vibration of the methylene group on C6 results in two bands in the region above 1450 cm^{-1} ; the methylene scissors vibration is

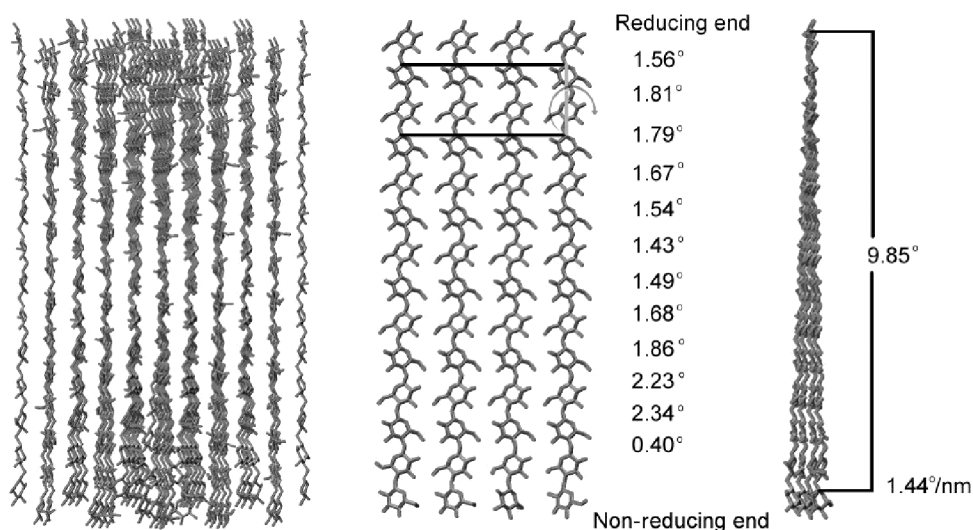


Figure 6.7 Views of the central portion of one center chain and one origin chain from the middle of the diagonal crystal, illustrating the inter-plane hydrogen bonds which can occur after the center chain primary alcohol groups rotate to the GG conformation. Hydrogen bonds between layers are indicated with dashed lines. (Reproduced in color as Plate 10.)

the only one that gives rise to a band above 1450 cm^{-1} . Thus, the occurrence of two bands is consistent with the presence of two rotamers.

In the GG conformation, the primary alcohol groups are essentially perpendicular to the average planes of the anhydroglucose rings and as a result are pointing up and down toward the origin chains of the layers above and below. In this conformation, the exocyclic groups can make good O6–O2 hydrogen bonds between layers. Since under normal conditions cellulose exhibits no tendency for layers to slip relative to one another, the existence of such stabilizing hydrogen bonds may not seem so implausible. However, in this conformation steric clashes between the center chain primary alcohol groups and the origin layers above and below force the center chains to tilt significantly with respect to the plane of their own layer. Such a tilt was also found by Heiner and Teleman (24).

Probably the most significant change that occurred during the simulations was that the aggregate quickly developed a small right-hand twist during the heating and equilibration interval and the twist remained relatively stable throughout the rest of the simulation. Figure 6.7 illustrates this twist, with the middle hydrogen-bonded sheet shown in detail. In this figure, the average twist angle for each successive glycosidic linkage is shown. These angles are defined as the dihedral angle for the four C1 carbon atoms illustrated as joined by the dark lines in the figure. Although this angle varies considerably near the non-reducing end, apparently because of edge effects, the twist in the middle of the chain is fairly constant at around $1.4\text{--}1.7^\circ$ per linkage, with an overall twist for this short oligosaccharide segment of almost 9.9° calculated from the first and last rows (which includes considerable irregularity due to the highly frayed structure of the non-reducing ends).

Imposition of the constraint of the symmetry of space group $P2_1$ confines cellulose chains to an exact twofold helix, and this constraint can be satisfied by many combinations of torsion angles across the glycosidic linkage (reported either as $\phi_H = H1-C1-O-C4'$ and $\psi_H = C1-O-C4'-H4'$ or as $\phi_O = O5-C1-O-C4'$ and $\psi_C = C1-O-C4'-C5'$). However, the line connecting twofold helical structures for cellobiose in ϕ, ψ space does not coincide with a free energy minimum (25, 26). Cellulose oligomers in solution are extended, but do not have a flat ribbon structure (27, 28). The preference of cellulose chains to adopt conformations away from a twofold helix is frustrated in a crystalline state by packing and hydrogen bonding requirements. Equilibrium organization of the aggregate has each of the individual interior chains departing slightly from the flat starting structure, on average forming a right-handed helix. The helix of each chain corresponds to the overall twist of fiber in a manner similar to the twist seen in protein β -sheets (29, 30).

In addition to the findings of the simulation studies, it is helpful to consider the source of the helical patterns at the level of the individual monomers in cellulose; it is now accepted that anhydrocellobiose is the repeat unit of structure in cellulose as it implicitly defines the glucosidic linkage as well. The results of the earliest conformational energy mappings available (31, 32) show that two energy minima associated with variations in dihedral angles of glycosidic linkage correspond to relatively small left- and right-handed departures from glycosidic linkage conformations consistent with twofold helical symmetry. More recent all-atom conformational energy maps for cellobiose exhibit the same qualitative topology (25). Local minima also represent values of dihedral angles very similar to those reported for cellobiose and methyl β -cellobioside on the basis of crystallographic analyses (9, 10). The relationship between different conformations is represented in Figure 6.8, which was adapted (6) from a diagram first presented by Reese and Skerrett (31).

Figure 6.8 is a ψ/ϕ map presenting different categories of information concerning conformation of the anhydrocellobiose unit as a function of the values of two dihedral angles

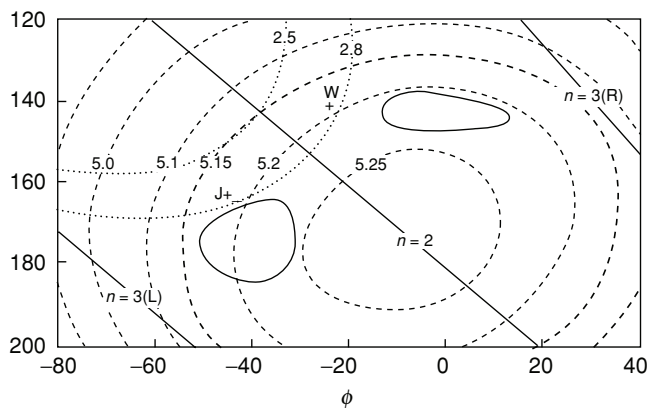


Figure 6.8 ψ/ϕ map adapted from Ref. (9). (---) Loci of structures with constant anhydroglucose repeat periods as noted in Angstroms. (...) Loci of structures of constant intramolecular bond $O5-O3^1$ distances. (—) Contours of potential energy minima based on non-bonded interactions in cellobiose. J – cellobiose; W – β -methylcellobioside; $n = 2$, the twofold helix line; $n = 3$, the threefold helix line; (R) right handed; (L) left handed. The Meyer and Misch structure is at $\psi = 180$, $\phi = 0$.

about bonds in the glycosidic linkage. ψ is defined as the dihedral angle about the bond between C4 and the glycosidic linkage oxygen and ϕ as the dihedral angle about the bond between C1 and the glycosidic linkage oxygen. The parallel lines indicated by $n = 3(L)$, 2, and 3(R) represent values of dihedral angles consistent with a left-handed threefold helical conformation, a twofold helical conformation, and a right-handed threefold helical conformation, respectively. A twofold helical conformation inherently does not have a handedness to it. Dashed contours represent conformations that have indicated repeat period per anhydroglucose unit; the innermost represents a period of 5.25 Å corresponding to 10.5 Å per anhydrocellobiose unit. Two dotted lines indicate conformations corresponding to values of 2.5 and 2.8 Å for the distance between the two oxygen atoms anchoring the intramolecular hydrogen bond between the C3 hydroxyl group of one anhydroglucose unit and the ring oxygen of the adjacent unit; values bracket the range wherein hydrogen bonds are regarded as strong.

The two domains defined by solid lines on either side of the twofold helix line ($n = 2$) represent the potential energy minima calculated by Reese and Skerrett for different conformations of cellobiose (31). Finally, the points marked by J and W represent the structure of cellobiose determined by Chu and Jeffrey (9) and that of methyl β cellobioside determined by Ham and Williams (10). The key point to be kept in mind with this diagram is that structures along the twofold helix line and with a repeat period of 10.3 Å per anhydrocellobiose unit possess an unacceptable degree of overlap between the van der Waals radii of the hydrogen atoms on either side of the glycosidic linkage.

Figure 6.8 shows that the structure of glycosidic linkage in cellulose is not likely to coincide with the line representing twofold helical structures. Rather, it is likely to be on either side of the twofold helix line as are the structures of cellobiose determined by Chu and Jeffrey, designated (J) and of β -methylcellobioside determined by Ham and Williams designated (W). However, because of the repeat distance per anhydroglucose unit, one would expect the glycosidic linkages in cellulose to be much closer to the twofold line than are the two dimeric structures. On the other hand, the SS ^{13}C NMR spectra show a splitting of the resonances at C1 and C4. Thus, it seems plausible that values of the glycosidic dihedrals in the cellulose chain might alternate between a small left-handed departure and a slightly larger right-handed departure from the twofold helix line. The net effect would be a slow, long-period, right-handed helical structure. Such an alternating pattern was observed in the stable equilibrium structure at the conclusion of the MD simulation (14). This pattern demonstrates that the long-period helical twist is a consequence of important characteristics of glycosidic linkages in cellulose rather than an artifact of a complex simulation.

From a broader perspective, a very important result of molecular modeling validates the approach represented by the theoretical model used. The finding that the cellulose aggregate is stable reflects that cellulose is insoluble in water beyond the octamer. The stability of the aggregate at equilibrium is not the result of any constraints or boundary conditions imposed on the solution of the equilibrium structure, but rather evidences that the molecular modeling has captured some essential distinctive properties of cellulose. Indeed, the measure of its true approximation of the nature of cellulose is that the insolubility is predicted for chains that are 12 anhydroglucose units in length. Furthermore, the results of the modeling are consistent with microscopic observations of long-period helical structures, and they explain the structure of the HCH scissors vibration bands in the Raman spectra of I_α and I_β .

The results are also consistent with the effects of small-diameter fibrillation to be discussed further below.

It is important at this point to return to Figure 6.3, panel B and consider its implications. The application of a long period of 1200 nm to all of the fibrils is intended to allow comparison of the effects of lateral dimensions on the twisting. A period of 1200 nm is evidently too short a period for fibrils of *Valonia* and *Halocynthia*, considering that such periods are rarely observed in electron microscopy. On the basis of the observation of a long period of 1200 nm for *Micrasterias*, we anticipate that the period for a 20 by 20 nm fibril is likely to be 2500 nm or more. That dimension is 2.5 μm and would be well beyond the field size in a high-magnification electron micrograph.

Another important point most obvious for the 20 by 20 nm fibril in panel B is that the center chain remains linear though it will be twisted by 90° . The other chains, however, develop some curvature so that the corner chains are obviously quite curved, and curvature increases with distance from the center. Thus, the resistance of the aggregate to inherent tendencies of the cellulose molecules to acquire a helical orientation increases with lateral dimension. The possibility of shear stresses developing within a fibril increases with lateral dimension also. This may well be why load-bearing structures of higher plants have fibrils with such small lateral dimensions. Because of small diameters, they are not likely to develop significant internal shear stresses. Because their association with neighboring fibrils is mediated by water, they can move parallel to each other when under load. Panel B in Figure 6.3 clearly suggests that as the lateral dimensions are reduced, the long-period helical twist can be more easily accommodated. The implications of the long-period twist for the subject of cell wall deconstruction will be discussed further in a following section.

6.3.2 Raman spectra

The Raman spectra of two forms of native celluloses, I_α and I_β , have been discussed in detail elsewhere (13), but are presented here in support of evidence for a helical model for native biological structures. Furthermore, these forms are presented to support findings from the molecular modeling program that two different rotamers of the exocyclic group at C6 occur in the native state. Finally, spectra are presented to illustrate the effects of lateral dimensions on spectral resolution.

The spectra presented in Figure 6.9 were recorded for samples of the alga *Valonia ventricosa* and the tunicate *Halocynthia roretzi*. The first is an alga wherein the fibrils of cellulose are 65% I_α and 35% I_β . The tunicate appears to be predominantly I_β . The importance of these spectra derives because fibrils in both instances are approximately 20 by 20 nm in lateral dimension. Thus, resolution of the spectra is sufficient to allow confident discussion of their interpretation. Two features of the spectra are noteworthy. The first supports the observation based on molecular modeling that two distinct rotamers of the exocyclic group at C6 occur in the native form, which is evident from the appearance of two distinct bands above 1450 cm^{-1} . Presently, it is not possible to associate the individual bands with the corresponding rotamer, but there is little question that two distinct rotamers occur. The relative intensities of the two bands differ in the spectra of the two forms. In the spectrum of the predominantly I_α *Valonia*, the lower frequency band is higher than it is in the spectrum of the predominantly I_β *Halocynthia*. The occurrence of two bands above 1450 cm^{-1} is

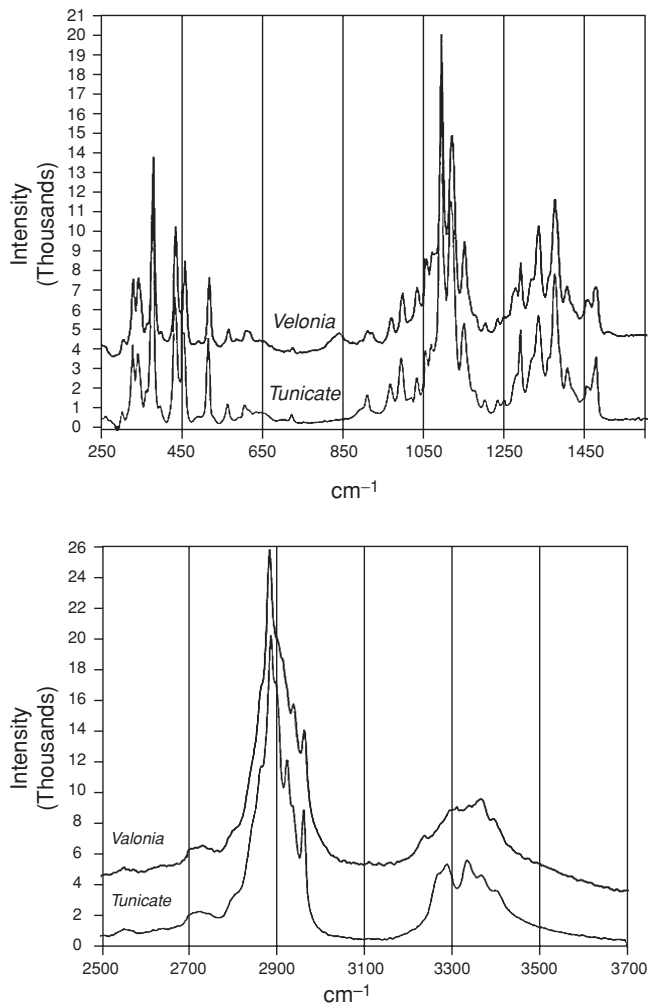


Figure 6.9 The Raman spectra of tunicate (*Halocynthia roretzi*) and *Valonia ventricosa* celluloses in the Raman-active fundamental regions.

unique to the two forms I_{α} and I_{β} . The spectra of celluloses **II** and **III** both have a single sharp band in this region (33).

Since questions have been raised within the cellulose science community regarding sensitivity of Raman spectra to molecular conformation, we present here some of the considerations that persuade us. Raman spectroscopy is a branch of vibrational spectroscopy complementary to infrared (IR). Raman spectra are no less sensitive to perturbations of molecular structure or environment than IR and are indeed better suited to studies of biological systems because of the very low scattering coefficient of water. Both IR and Raman spectroscopy involve transitions between molecular vibrational states. The key difference is that Raman spectra are more sensitive to vibrational transitions involving highly covalent

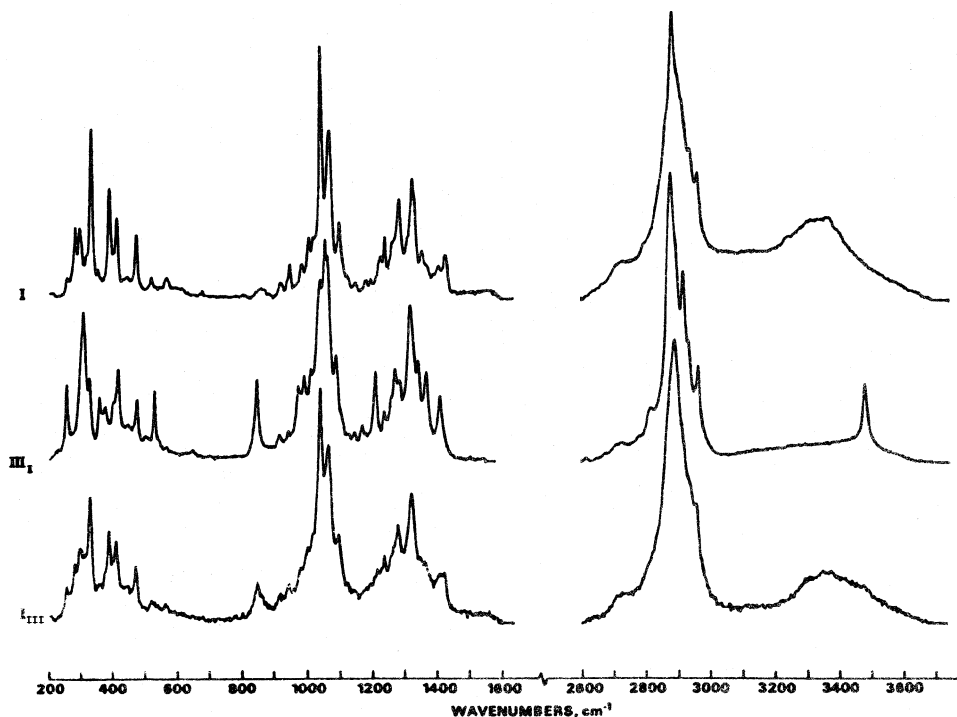


Figure 6.10 Raman spectra of *Cladophora glomerata* in its native state **I**, after conversion to cellulose **III** in liquid ammonia (**III_I**) and after recovery in the **I** form by boiling in water (**I_{III}**) [Ref. (12)].

bond systems, whereas IR spectra are more sensitive to transitions involving highly polar systems of bonds.

The basis for establishing sensitivity of Raman spectra to molecular conformation was established through extensive normal-coordinate analyses of six classes of model compounds related to saccharides (34–42). The most persuasive are analyses of vibrational spectra of the inositols (34, 39). The inositols are cyclohexane hexols, differing from each other only in the distribution of hydroxyl groups between axial and equatorial orientation and their positions relative to each other. Differences in their spectra leave very little doubt that for molecules where the skeletal structure is made up of C–C and C–O bonds, for which reduced masses, bond energies, and force constants are similar and the coupling of vibrational modes is high, the individual spectra are determined by organization of atoms in space relative to each other within the molecule.

For demonstration of the sensitivity of Raman spectra to conformation particularly in the context of celluloses, we present in Figure 6.10, Raman spectra of cellulose from *Cladophora glomerata*, which is a fresh water alga that has fibrils very similar to those of *Valonia*, both with respect to cross section and the balance between I_α and I_β . Spectrum (**I**) is for the alga in its native state after purification with an acid chlorite treatment; it is very similar to the spectrum of *Valonia* presented in Figure 6.9. Spectrum (**III_I**) is recorded after conversion to cellulose **III** by treatment in anhydrous liquid ammonia at -30°C ; the very high level of order is retained by allowing the ammonia to evaporate gradually at ambient temperature.

Spectrum (I_{III}) is after recovery of cellulose **I** primarily in the I_{β} form from cellulose **III** by boiling in water.

Three comparisons in Figure 6.10 are noteworthy. First are the significant differences between the spectra of native cellulose **I** and **III**_I; the key difference between celluloses **I** and **III** are differences in conformation. Second, the spectrum of I_{III} recovers similarity to that of the alga but is now less well resolved because of fibrillation accompanying the conversion back to cellulose **I** first noted by Chanzy and coworkers (43). Lateral dimensions of fibrils of cellulose I_{III} range between 3 and 6 nm, which are in contrast to the lateral dimensions of the original *Caldophera* cellulose at 20 by 20 nm. The spectrum of I_{III} is almost indistinguishable from that of cotton. Thus, lateral dimensions of nanofibrils are clearly very important to resolution of bands in the spectra. Most of the bands that are very sharp with relatively low bandwidth for cellulose **I** from the native *Caldophera* with 20 by 20 nm fibrils are now broadened considerably in the spectrum of I_{III} . Chanzy and coworkers found a similar effect with SS ¹³C NMR spectra. Finally, the Raman spectrum of tunicate I_{β} cellulose shown in Figure 6.9 is much more similar to those of the I_{α} algal celluloses in Figures 6.9 and 6.10, than to that of the fibrillated I_{β} sample of I_{III} .

We conclude that spectra in Figure 6.9 confirm that conformations of the I_{α} and I_{β} forms are almost identical and that results of the diffractometric studies reflect two erroneous assumptions. The first, by Atalla and VanderHart that I_{α} and I_{β} are “two distinct *crystalline* forms” rather than simply “two distinct forms,” the second by the authors of diffractometric studies in further assuming that I_{α} and I_{β} belong to different crystallographic space groups, which is also in contrast to the findings of Sugiyama and coworkers (12).

6.3.2.1 Native celluloses in living plants are not crystalline in the classical sense

This echoes the observation by Cross and Bevan almost a century ago when crystallinity of cellulose was first proposed, “The root idea of crystallography is identical invariability while the root idea of the world of living matter is essential individual variation” (44). Recognizing the species and tissue specificities of the structures of native celluloses is essential to progress in understanding the diversity of cellulose synthases encoded in the genomes of plants and to understanding the even greater diversity of the cellulases produced by species-specific plant pathogens.

6.3.3 Alternative patterns of aggregation

The above considerations leave little doubt that the primary aggregates of cellulose emerging from individual rosettes are likely to have a long-period helical character. Depending on the cooperative association of rosettes during biogenesis and their relative mobility within the plasma membrane, these primary aggregates will come together to form a secondary aggregate that may vary in relative organization and have a longer period. At this and higher levels of assembly of native celluloses, tissue and species specificities are expected to arise. One of the key determinants will be the degree to which the synthase rosettes act cooperatively; we anticipate that this is one key point of entry of distinctive genomic information. The variability of higher levels of aggregation is illustrated in Figure 6.2, where it is obvious that

individual fibrils often occur in pairs or triads, and these in turn can be intimately integrated in higher level associations into aggregates of multiple fibrils.

In developing a foundation for experimental studies of native celluloses that explore the relationship between structure and genomic information, it is helpful to consider alternative patterns of aggregation and assess whether they might be altered during isolation from higher plant tissues, whether for experimental studies or in industrial processes that use different celluloses as feedstocks. To accomplish this, a number of models have been represented in the same manner as was done in Figure 6.3, in order to explore how the different nanofibrils with helical structures might come together in the aggregation to form the next higher or secondary level of nanostructure.

To facilitate visualization of factors that enter into aggregation of nanofibrils, Figure 6.11 was developed. In panel A, a 6 by 6 nm nanofibril is represented both as a single nanofibril

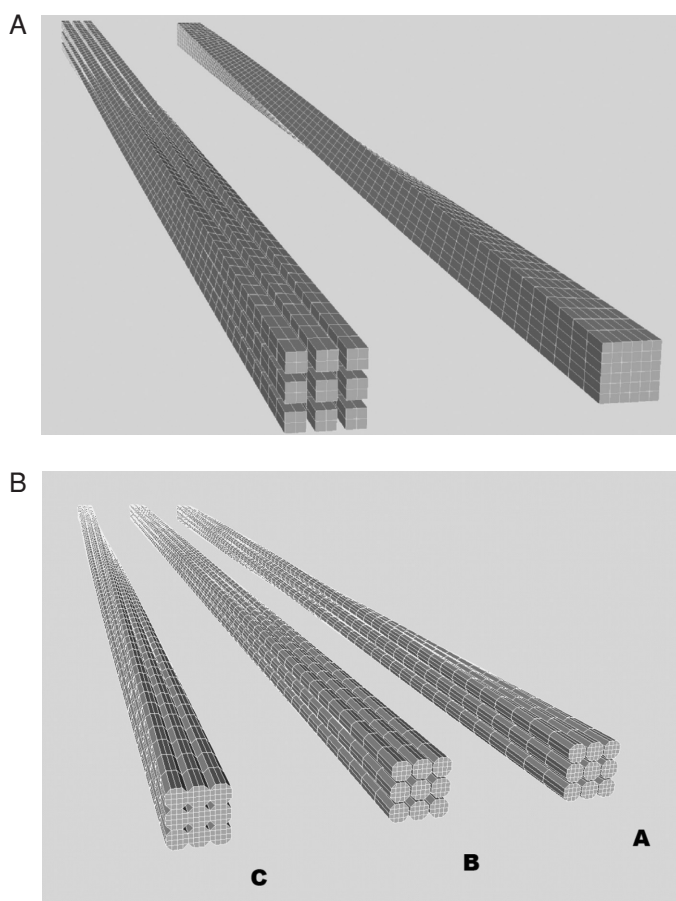


Figure 6.11 Effects of packing. Panel A compares the helical form of a 6 by 6 nm nanofibril with that of nine 2 by 2 helical nanofibrils packed as close as possible without surface intersections. (Reproduced in color as Plate 11.) Panel B shows helical nanofibrils assembled in different patterns. (Reproduced in color as Plate 12.)

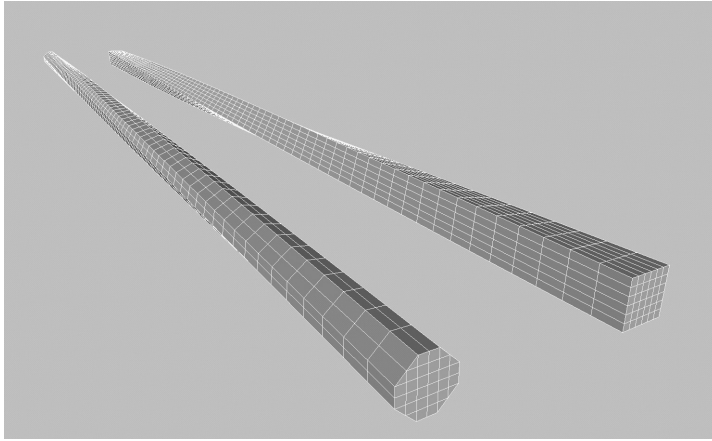


Figure 6.12 Aggregation into a larger nanofibril. (Reproduced in color as Plate 13.)

with the specified helical period and then as an assembly of nine 2 by 2 helical nanofibrils of the same period. The 2 by 2 nm fibrils were considered representative of the most elementary nanofibrils formed. We recognize at the outset that while the square cross section may be stable at the 20 by 20 nm level, it is not likely to be stable at the 2 by 2 or 6 by 6 levels. Surface phenomena would lead to their transformation to polygonal cross sections of higher order. However, we believe panel A to be a helpful intermediate representation because the helical pattern is more clearly visualized.

To improve the approximation to reality, the aggregates in panel B were constructed. First, the corners of the 2 by 2 fibrils in panel A were removed so that the cross section becomes octagonal and can more closely approximate the circular or ellipsoidal polygonal cross sections usually observed for higher plants. The fibrils were then assembled in three different modes. In A, the fibrils were twisted individually and then the assembly also twisted. In B, the individual fibrils were each subjected to a twist of 90° over the 300-nm period, and then packed as closely as possible without their surfaces intersecting. Finally in C, the fibrils were collectively subjected to the twist. These of course represent the most simply visualized members of secondary patterns of aggregation, and other patterns are expected to occur also as a result of variability in patterns of cooperative associations of synthase rosettes. A number of circumstances can be envisioned for the further aggregation of the nanofibrils. In the pattern represented by C at the left of panel B, nanofibrils retain a coherence of order relative to each other that might allow them to come together to form a fibril similar to the single fibril in panel A, but with the corners rounded off. This case is illustrated in detail in Figure 6.12.

The type of aggregation shown in Figure 6.12 may in fact be responsible for formation of the types of fibrils that occur when cellulose is deposited alone in the cell wall, as in nanofibrils of cotton or ramie. The same would apply to aggregation of elementary nanofibrils in algae, though their assembly processes differ from those of higher plants. Pattern B of Figure 6.12 would not be mechanically stable, so we believe this pattern is unlikely to occur

in load-bearing tissues. It may occur in other contexts where its distinctive character fits selected functions in the cell wall. The pattern A shown on the right side of panel B in Figure 6.11 is the most likely pattern of aggregation when cellulose is deposited in the presence of other cell wall constituents that might influence the progress of the aggregation.

When one considers flexibility of the fibrils and the possible nearest neighbor, lateral interactions between them, it is not possible to anticipate which pattern of aggregation occurs in any particular plant tissue. And it is at this level of aggregation that we believe the balance between the I_α and the I_β forms is established. When the most elementary nanofibrils co-aggregate with adjacent ones, one can expect the patterns of hydrogen bonding to be modified. It seems very likely that the inherent self-assembly characteristic of the cellulose molecule in its native conformation is more a function of its skeletal organization. The hydrogen bonding patterns appear to be secondary determinants of the aggregation. This, of course, is also consistent with findings from the molecular modeling program.

From a mechanical point of view, we believe the most likely pattern in load-bearing tissues of higher plants is pattern A on the right in panel B of Figure 6.11. This view is influenced, in part, by the pattern being the closest approximation to patterns used at the macroscopic level in the design and construction of cables and ropes. This pattern is likely the most efficient load-bearing structure. These visualizations are derived from construction of mathematical models and are not artistic depictions. However, we recognize that there are other patterns of secondary aggregation.

6.3.3.1 *Changes during isolation*

With the patterns of aggregation indicated for native celluloses in living plants, two stages in the isolation processes will influence the final pattern of aggregation. The first is elevation of temperature during most such processes. The second is the effect of drying the sample as the last stage in the processes of isolation of native celluloses.

In the methods most often used for purification of celluloses for investigation, two methodologies stand out. For tissues that are relatively pure celluloses, boiling in dilute caustic under nitrogen is a very common practice. And the mildest procedure used for isolating cellulose from highly lignified woody tissue begins with the acid chlorite method, usually carried out at 70°C. Both of these methods involve temperature elevation to levels that can alter the patterns of aggregation in ways that make the cellulose more recalcitrant during hydrolysis.

Though of course we know that cellulose oligomers are essentially insoluble in water at the octamer and beyond, we expect it to be hydrated in its native state at the level of the elementary nanofibrils or at least at the first level of aggregation. We also know that all cellulose derivatives that do not have any ionic substitution have negative correlations of solubility with temperature (45), with most of them precipitating between 40 and 60°C. This is likely also true of cellulose at the nanoscale level. So, temperature elevation is likely to change the state of aggregation of native celluloses. The only path to avoiding this effect is isolation at ambient temperatures. Such procedures are possible, though they require patience.

Another important factor is that celluloses are deformed during drying procedures. The more common methods are drying samples in an oven at 105°C or freeze-drying them. The first hint of distortion during the drying process came during the early solid state ^{13}C NMR studies, where researchers observed that moistening samples to a level of 20% moisture

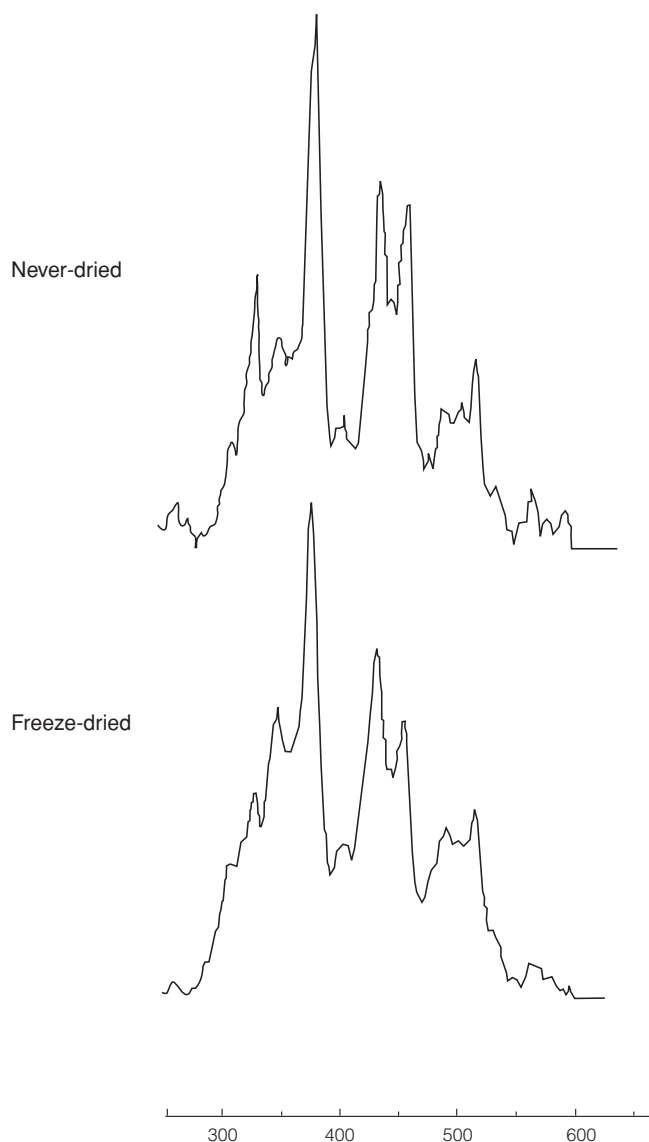


Figure 6.13 Raman spectra of bleached spruce kraft pulp, before and after freeze drying.

resulted in enhancement of the resolution of spectra. It was thought that this effect could be due to enhancement of cross-polarization efficiency or distortions during the drying process that are relaxed by the added moisture. Definitive answers to this question on the basis of SS ^{13}C NMR were not pursued (5).

However, in a retrospective review of Raman spectra recorded over many years in Atalla's laboratory, one set of spectra stood out as clear evidence of the occurrence of such distortions during dehydration. The spectra are shown in Figure 6.13. They are spectra of a

sample of never-dried bleached spruce kraft pulp recorded before and after freeze drying, thus avoiding further effects of temperature. In the course of the kraft-pulping process, the fibers had been subjected to temperatures in the range of 170–180°C, which resulted in a high degree of aggregation of cellulose well beyond the level in its native state. In spite of that, an effect of freeze drying is observed.

The upper spectrum of the never-dried sample bands are broadened dramatically as a result of the drying process; the spectrum has many weak bands that appear to be broadened and drowned into the background after freeze drying. These effects suggest that the freeze-drying process, by removing too much of the water needed to lubricate the motion of the nanofibrils relative to each other, results in distortions of the nanofibrils as they hydrogen-bond with each other rather than remaining isolated in a vacuum.

From these observations, we conclude that almost all studies of structures of cellulose undertaken so far have been of celluloses that have been modified to varying degrees in the course of isolation. In addition to the evidence presented here, X-ray diffraction studies and solid state ^{13}C NMR measurements on cotton fibers have shown that cotton fibers in the unopened boll tend to decline in degree of order upon opening of the boll and subsequent dehydration of the cotton fibers; the X-ray diffraction patterns are broadened and the spectra loose resolution. This can only be interpreted as resulting from some distortion of the native order as the native moisture is removed.

More recently, we have observed in our laboratory that bacterial celluloses manifest changes in their spectra after dehydration. The changes are not dramatic, but they reveal a broadening of the bands that are more sharply resolved in the never-dried samples. Samples for the spectra in Figure 6.9 were treated with acid chlorite at ambient temperature. The samples used in all of the prior studies of structure (10, 11, 46) were boiled in NaOH or KOH solutions for extended periods, followed by extended periods of exposure to very concentrated sulfuric acid, presumably to remove the “amorphous fractions.” It is useful to consider the effect of boiling on an aggregate of cellulose nanofibrils, given the effects of temperature on cellulose hydration noted above. Figure 6.14 shows stages of further aggregation likely to occur at elevated temperatures in native higher-plant celluloses. The secondary aggregate shown as A in panel B of Figure 6.11 has been subjected to tighter compaction to varying degrees.

Thus, a previously uniform aggregate with low curvature is forced into a form where it has extended parallel primary structures, interrupted by a highly twisted connection; this is not unlike the nanofibril of *Micrasterius denticulata* shown in the left panel of Figure 6.1. The highly twisted region is then regarded as amorphous. A previously uniform helical natural structure that has a distinctive long period and limited curvature has been transformed by thermal dehydration into an approximation of the fringed micelle model introduced in the first half of the last century. It is our view that the effect illustrated in Figure 6.14 is an inevitable consequence of boiling in caustic solutions for extended periods.

Since our primary concern in this volume is cellulose in higher plants, we have developed another model structure intended to approximate fibrils in higher plants. It consists of four elementary fibrils in a 2 by 2 arrangement. The elementary fibrils are approximately 4 nm in diameter. These are shown in Figure 6.15 where now the long period is represented as 300 nm, and the segment shown is 75 nm long and undergoes a twist of 90°. This approximation of the native form is represented in pattern A as it is expected to occur in the native state in the living plant. The effect of dehydration, taken to be tight aggregation resulting in

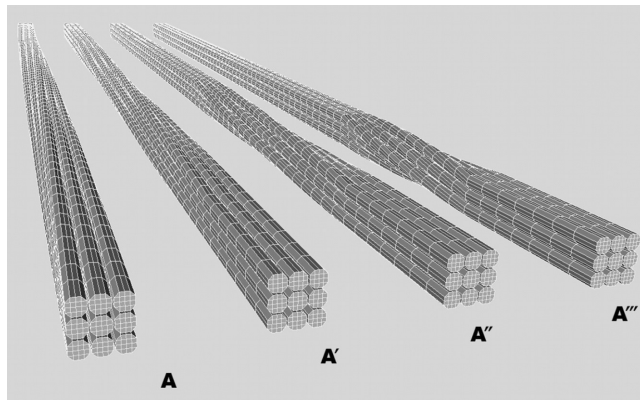


Figure 6.14 The secondary aggregate A is identical to the one designated as A on the right side of panel B in Figure 6.11, thus it has the primary aggregates twisted to match a 1200-nm period helix and then the secondary aggregate is subjected to a similar helical deformation; this is the uniform secondary aggregate proposed as the most efficient load-bearing structure. In the progression toward A', A'', and A''', different levels of further aggregation are depicted such that portions of the primary aggregates are made parallel and thus more closely approximate a lattice. In A', it is 30%, then 65% in A'', and finally 85% in A'''.

approximately 85% conversion to the more tightly aggregated form, is shown in pattern A' of the figure.

Thus, a consequence of the dehydration, whether by drying or by temperature elevation, is the creation of more tightly aggregated domains that are likely to be more recalcitrant to hydrolytic action, whether by enzymes or acids. Furthermore, we suggest that the *linear*

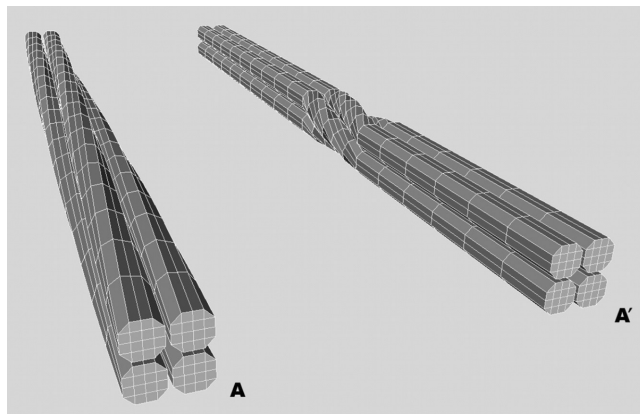


Figure 6.15 Representations of the patterns of aggregation in a fibril 4 nm in diameter, consisting of four elementary fibrils. In A, each of the elementary fibrils has a helical twist with long period of 300 nm and the collective aggregate also has a helical twist with a long period of 300 nm. In A', 85% of the length has been forced to align in a parallel pattern resulting in what can appear to be nanocrystalline domains.

parallel segments, which are artifacts of isolation processes are easily mistaken for naturally occurring crystalline domains.

6.4 Alternative approaches to the problem of crystallinity

In light of the observations discussed above, it is important to consider how the new understanding and insights into the phenomenology of native celluloses can lead to process improvements that can enhance the digestibility of native celluloses to glucose.

Two alternative approaches come to mind. The first is to develop improved pretreatment processes that avoid the tight aggregation associated with dehydration. Given the factors listed above, it would appear that the only way to avoid this type of aggregation is to undertake the pretreatment at temperatures below 40–50°C. Although at present no processes are known that can be readily scaled up for application on an industrial scale, this is an arena that has received very little attention, in part, because it has always been assumed that the tightly aggregated “crystalline” domains are pre-existing constituents in the cell walls of living plants. Some laboratory scale methods are indeed available for chemical deconstruction at ambient temperatures and others have recently been developed. Exploration of the possibility of adapting these methods for industrial scale is a logical arena for further investigation.

The second alternative is to continue the application of elevated temperatures in order to accelerate the rates of removal of “encrustants,” but to follow the pretreatment with processes that can reverse the tight aggregation occurring during the pretreatment. Although such processes have been known for some time, they have seemed of academic interest only. It is very likely that some of these methods can be scaled up for industrial utilization.

References

1. Preston, R.D. (1974) *The Physical Biology of Plant Cell Walls*. Chapman and Hall, London.
2. Frey-Wyssling, A. (1976) *The Plant Cell Wall*. Gebruder Borntraeger, Berlin.
3. Atalla, R.H. (1999) The celluloses. In: *Comprehensive Natural Products Chemistry*, Vol. 3 (ed. M. Pinto). Elsevier, New York.
4. Atalla, R.H. & VanderHart, D.L. (1984) Native cellulose: A composite of two distinct crystalline forms. *Science*, **223**, 283.
5. VanderHart, D.L. & Atalla, R.H. (1984) Studies of microstructure in native celluloses using solid state ^{13}C NMR. *Macromolecules*, **17**, 1465.
6. Atalla, R.H. (1979) Conformational effects in the hydrolysis of cellulose. *Advanced Chemistry Series*, **181**, 55.
7. Nishiyama, Y., Sugiyama, J., Chanzy, H. & Langan, P. (2003) Crystal structure and hydrogen bonding system in cellulose I_α from synchrotron X-ray and neutron fiber diffraction. *JACS*, **125**, 14300.
8. Nishiyama, Y., Chanzy, H. & Langan, P. (2002) Crystal structure and hydrogen-bonding system in cellulose I_β from synchrotron X-ray and neutron fiber diffraction. *JACS*, **124**, 9074.
9. Chu, S.S.C. & Jeffrey, G.A. (1968) The refinement of the crystal structures of β -D-glucose and cellobiose. *Acta Crystallography*, **B24**, 830.
10. Ham, J.T. & Williams, D.G. (1970) The crystal and molecular structure of methyl β -cellobioside-methanol. *Acta Crystallography*, **B29**, 1373.

11. Honjo, G. & Watanabe, M. (1958) Examination of cellulose fibre by the low-temperature specimen method of electron diffraction and electron microscopy. *Nature*, **81**, 326.
12. Sugiyama, J., Harada, H., Fujiyoshi, Y. & Uyeda, N. (1985) Lattice images from ultrathin sections of cellulose microfibrils in the cell wall of *Valonia macrophysa* Kutz. *Planta*, **166**, 161.
13. Atalla, R. (in press) Nanoscale structures in native celluloses, their species and tissue specificities, and their denaturation during isolation. *Carbohydrate Research*.
14. Matthews, J.F., Skopec, C.E., Mason, P.E., Zuccato, P., Torget, R.W., Sugiyama, J., Himmel, M.E. & Brady, J.W. (2006) Computer simulation studies of microcrystalline cellulose I_β. *Carbohydrate Research*, **341**, 138–152.
15. Hanley, S.J., Revol, J.-F., Godbout, L. & Grey, D.G. (1997) Atomic force microscopy and transmission electron microscopy of cellulose from *Micrasterias denticulate*: Evidence for a chiral helical microfibril twist. *Cellulose*, **4**, 209.
16. Haigler, C.H. (1991) Biosynthesis of bacterial celluloses. In: *Biosynthesis and Biodegradation of Cellulose* (eds. C.H. Haigler & P.J. Weimer), p. 99. Marcel Dekker, New York.
17. Hirai, A., Tsuji, M. & Horii, F. (1998) Helical sense of ribbon assemblies and splayed microfibrils of bacterial cellulose. *Sen'i Gakkaishi*, **54**, 506.
18. Ding, S.-Y. & Himmel, M.E. (2006) The maize primary cell wall microfibril: A new model derived from direct visualization. *Journal of Agricultural and Food Chemistry*, **54**, 597.
19. Tadokoro, H. (1979) *Structure of Crystalline Polymers*, p. 6. Wiley, New York.
20. Marchessault, R.H. & Pérez, S. (1979) Conformations of the hydroxymethyl group in crystalline aldohexopyranoses. *Biopolymers*, **18**, 2369–2374.
21. Viëtor, R.J., Newman, R.H., Ha, M.-A., Apperley, D.C. & Jarvis, M.C. (2002) Conformational features of crystal-surface cellulose from higher plants. *The Plant Journal*, **30**, 721–731.
22. Newman, R.H. & Davidson, T.C. (2004) Molecular conformations at the cellulose–water interface. *Cellulose*, **11**, 23–32.
23. Sturcová, A., His, I., Apperley, D.C., Sugiyama, J. & Jarvis, M.C. (2004) Structural details of crystalline cellulose from higher plants. *Biomacromolecules*, **5**, 1333–1339.
24. Heiner, A.P. & Teleman, O. (1997) Interface between monoclinic crystalline cellulose and water: Breakdown of the odd/even duplicity. *Langmuir*, **13**, 511–518.
25. Kuttel, M.M. (2003) *Simulations of Carbohydrate Conformational Dynamics and Thermodynamics*. PhD dissertation, Department of Chemistry, University of Cape Town, Rondebosch, South Africa.
26. French, A.D. & Johnson, G.P. (2004) What crystals of small analogs are trying to tell us about cellulose structure. *Cellulose*, **11**, 5–22.
27. Almond, A. & Sheehan, J.K. (2003) Predicting the molecular shape of polysaccharides from dynamic interactions with water. *Glycobiology*, **13**, 255–264.
28. Sugiyama, H., Hisamichi, K., Usui, T., Sakai, K. & Ishiyama, J.-I. (2000) A study of the conformation of β-1,4-linked glucose oligomers, cellobiose to cellobiose, in solution. *Journal of Molecular Structure*, **556**, 173–177.
29. Chothia, C. (1973) Conformation of twisted β-pleated sheets in proteins. *Journal of Molecular Biology*, **75**, 295–302.
30. Ho, B.K. & Curmi, P.M.G. (2002) Twist and shear in β-sheets and β-ribbons. *Journal of Molecular Biology*, **317**, 291–308.
31. Reese, D.A. & Skerrett, R.J. (1968) Conformational analysis of cellobiose, cellulose, and xylan. *Carbohydrate Research*, **7**, 334.
32. Melberg, S. & Rasmussen, K. (1979) Conformations of disaccharides by empirical force-field calculations: Part II, β-cellobiose. *Carbohydrate Research*, **71**, 25.
33. Atalla, R.H. & VanderHart, D.L. (1989) Studies of the structure of cellulose using Raman spectroscopy and solid state ¹³C NMR. In: *Cellulose and Wood: Chemistry and Technology* (ed. C. Schuerch), p. 169. Wiley-Interscience, New York.

34. Williams, R.M. & Atalla, R.H. (1984) Vibrational spectra of the inositols. *Journal of Physical Chemistry*, **88**, 508.
35. Pitzner, L.J. (1973) *The Vibrational Spectra of the 1,5-anhydropentitols*. PhD dissertation, The Institute of Paper Chemistry, Appleton, WI.
36. Pitzner, L.J. & Atalla, R.H. (1975) The vibrational spectra of the 1,5-anhydropentitols. *Spectrochimica Acta*, **31A**, 911.
37. Watson, G.M. (1974) *The Vibrational Spectra of the Pentitols and Erythritol*. PhD dissertation, The Institute of Paper Chemistry, Appleton, WI.
38. Edwards, S.L. (1976) *The Vibrational Spectra of the Pentose Sugars*. PhD dissertation, The Institute of Paper Chemistry, Appleton, WI.
39. Williams, R.M. (1977) *The Vibrational Spectra of the Inositols*. PhD dissertation, The Institute of Paper Chemistry, Appleton, WI.
40. Wells, H.A. (1977) *The Vibrational Spectra of Glucose, Galactose and Mannose*. PhD dissertation, The Institute of Paper Chemistry, Appleton, WI.
41. Wells, H.A. & Atalla, R.H. (1990) An investigation of the vibrational spectra of glucose, galactose, and mannose. *Journal of Molecular Structure*, **224**, 385.
42. Carlson, K.P. (1978) *The Vibrational Spectra of the Cellodextrins*. PhD dissertation, The Institute of Paper Chemistry, Appleton, WI.
43. Chanzy, H., Henrisat, B., Vincendon, M., Tanner, S.F. & Belton, P.S. (1987) Solid-state ^{13}C -NMR and electron microscopy study on the reversible cellulose I \rightarrow cellulose III_I transformation in Valonia. *Carbohydrate Research*, **160**, 1.
44. Cross, C.F. & Bevan, E.J. (1912) *Researches on Cellulose (III)*. Longmans, Green & Co, London.
45. Klug, E.D. (1971) Some properties of water-soluble hydroxyalkyl celluloses and their derivatives. *Journal of Polymer Science, Part C – Polymer Symposia*, **36**, 491–508.
46. Kono, H., Erata, T. & Takai, M. (2003) Determination of the through-bond carbon–carbon and carbon–proton connectivities of the native celluloses in the solid state. *Macromolecules*, **36**, 5131–5138.

Chapter 7

Lignins: A Twenty-First Century Challenge

*Laurence B. Davin, Ann M. Patten,
Michaël Jourdes, and Norman G. Lewis*

7.1 Lignin: molecular basis and role in plant adaptation to land

Life, as humanity understands it, has inextricably been tied since eons past to the successful evolutionary adaptation of aquatic flora to terrestrial environments. This conquest apparently first began with emergence of various “primitive” forms of, and/or forerunners to, tracheids (water-conducting elements) in the land-based plants during the late Ordovician to Silurian periods (>400 million of years) (1–3). Such land plant forms gradually became capable of efficient hydration and metabolism under “water-limited” conditions and hence attained an inherent ability to survive in a wide variety of habitats. Eventually, various other modified forms of plant cell walls also evolved, these containing – at least for the vascular apparatus – celluloses, hemicelluloses, lignins, and small amounts of proteins, as their main chemical/structural constituents. Ultimately, lignification provided the means by which large upright vascular plant forms could be produced, and which enabled some species to more successfully compete for photosynthetic energy. This, in turn, gave a molecular or structural basis for much of the plant biodiversity that humanity enjoys today in its many resplendent forms. Furthermore, in addition to competition for light, an upright growth habit (provided by a true vascular system), allowed for better spore/pollen dispersal, increasing the genetic variability and species range. Yet today, our knowledge of how plant cell wall assembly occurs is at the most rudimentary level.

This chapter focuses upon the lignins. Next to cellulose, they are Nature’s second most abundant organic substances, and are products of the phenylpropanoid pathway (Figure 7.1). Significantly, many of the monolignol/lignin-forming pathway steps apparently also evolved during transition of plants to the land habitat, providing broad adaptive advantage to some 350 000 or so distinct present-day vascular plant species (4). To put this pathway into the broader context of carbon allocation from photosynthesis, woody gymnosperm stems generally have lignin contents ~28–30%, whereas those in woody angiosperms are of lower amount (~20–24%) (5). In both cases, this represents a significant commitment and designation of the total carbon taken up during photosynthesis; indeed, the lignins are some of the most metabolically “expensive” of all plant products formed (6).

Studies of many different plant species (i.e., from gymnosperms to angiosperms) have established that their lignins proper, while often varying in monomeric compositions, are derived from the three monolignols **1**, **3**, and **5** (Figure 7.1); additionally, various grasses

contain small amounts (~10% or so) of *p*-hydroxycinnamate esters **30–32** linked to the monolignols (Figure 7.2A). Lignins are also generally designated as *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S), based on their aromatic ring substitution pattern (Figure 7.2B). Regardless of monomeric compositions, lignins are considered to be complex, amorphous, phenylpropanoid polymers within the lignocellulosic matrices of plant cell walls. This highly conserved monolignol deployment occurs irrespective of the remarkable differences encountered in both diverse body plans and architectures of the plant forms in existence today.

In spite of their overall abundance, the lignin biopolymers are quite challenging to work with. This is because of their aromatic and hence very hydrophobic character; their relative intractability due to difficulties to efficiently solubilize and characterize derivatives thereof and then only under quite drastic chemical degradation conditions (7); and their striking abilities to self-associate (8, 9) due to the strong electronic stabilization energies between the various subunits in the adjacent biopolymeric chains. In terms of their molecular architecture, the lignins are considered to be products of phenoxy-radical-derived coupling reactions as originally proposed by Erdtman in 1933 (10) in model studies using isoeugenol (**33**, Figure 7.2C); Figure 7.2D also illustrates several of the most abundant interunit linkage types known in lignins proper that can be at least partially quantified (discussed below).

It needs to be emphasized that the artistic depictions of lignin macromolecular configuration have changed rather enormously over a time span of nearly five decades, i.e., from originally being envisaged as a complex three-dimensional (Bakelite-like) phenolic polymer (11, 12) (Figure 7.3A), to that amenable for a computer-programmable simulation (Figure 7.3B) (13, 14), to that now more recently of essentially more linear macromolecular entities (e.g., Figure 7.3C) (15). All of these models though represent simply artistic and consequently quite artificial depictions of native lignin structure(s). Indeed, several quite abundant substructures (e.g., 8–1' or their forerunners, dibenzodioxocin, etc.) are even absent in Figures 7.3A and 7.3B; moreover, the very changing nature of such depictions underscores the fact that little truly systematic research has yet been carried out to investigate (and develop methodologies for) the study of lignin primary structure(s). Furthermore, there may be no other biopolymeric entities whose structure(s) has (have) been depicted in such quite arbitrary ways, particularly since experiments had neither been conducted nor devised to investigate both the actual biochemical mode of assembly and the structure(s) so obtained. Yet this has been the situation for lignins for almost a century now.

Interestingly, in the 1970s and 1980s, there was also much enthusiasm in identification of lignin-degrading enzymes, with various laccases and peroxidases (specifically manganese and lignin peroxidases) reported as the main degradative enzymes involved (16–19). Today, it is doubtful that either class has such a function, and neither has fulfilled the optimistic promise for industrial application once envisaged over two decades ago (20). Moreover, our understanding today of how lignin deconstruction (so-called biodegradation) occurs *in vivo* is still at a most rudimentary level. For this reason, research activities are now being directed and/or initiated toward identifying the nature of true lignin degrading enzymes, e.g., putative lignin depolymerases acting on specific linkage types (21).

Nevertheless, the fantastic diversity of the extant vascular plant species – in terms of not only their remarkable differences in size, shape, and growth rates, but also as to whether they have woody or non-woody character, etc. – has all depended upon the successful formation of the lignified vascular apparatus.

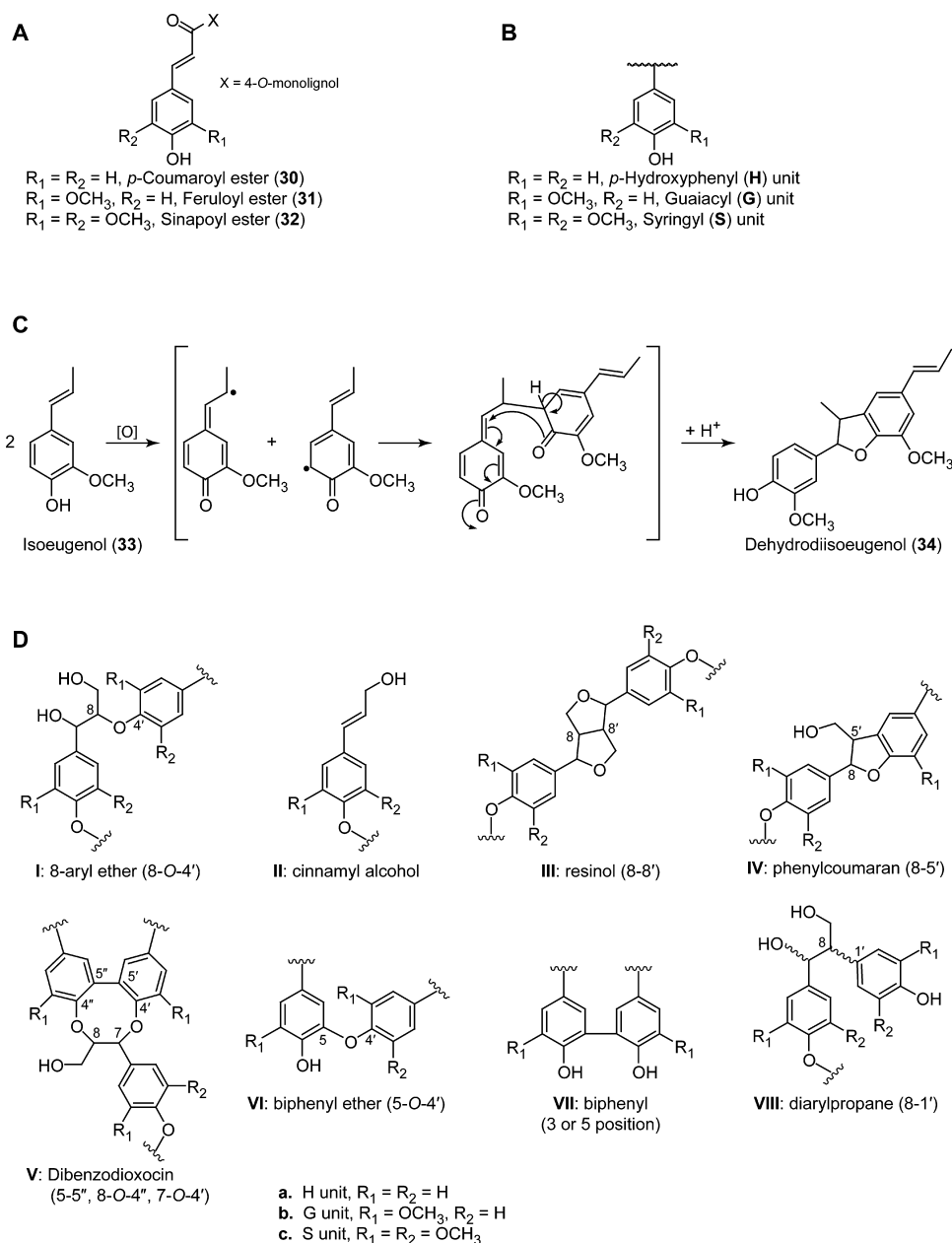


Figure 7.2 Phenylpropanoids. (A) Monolignol esters **30–32** found in grasses. (B) The aromatic ring components of lignins. (C) Erdtman's dehydrogenative coupling of isoeugenol (**33**), the basis of the lignin dehydrogenative polymerization model for monolignols. (D) Several dominant substructures present in native lignins.

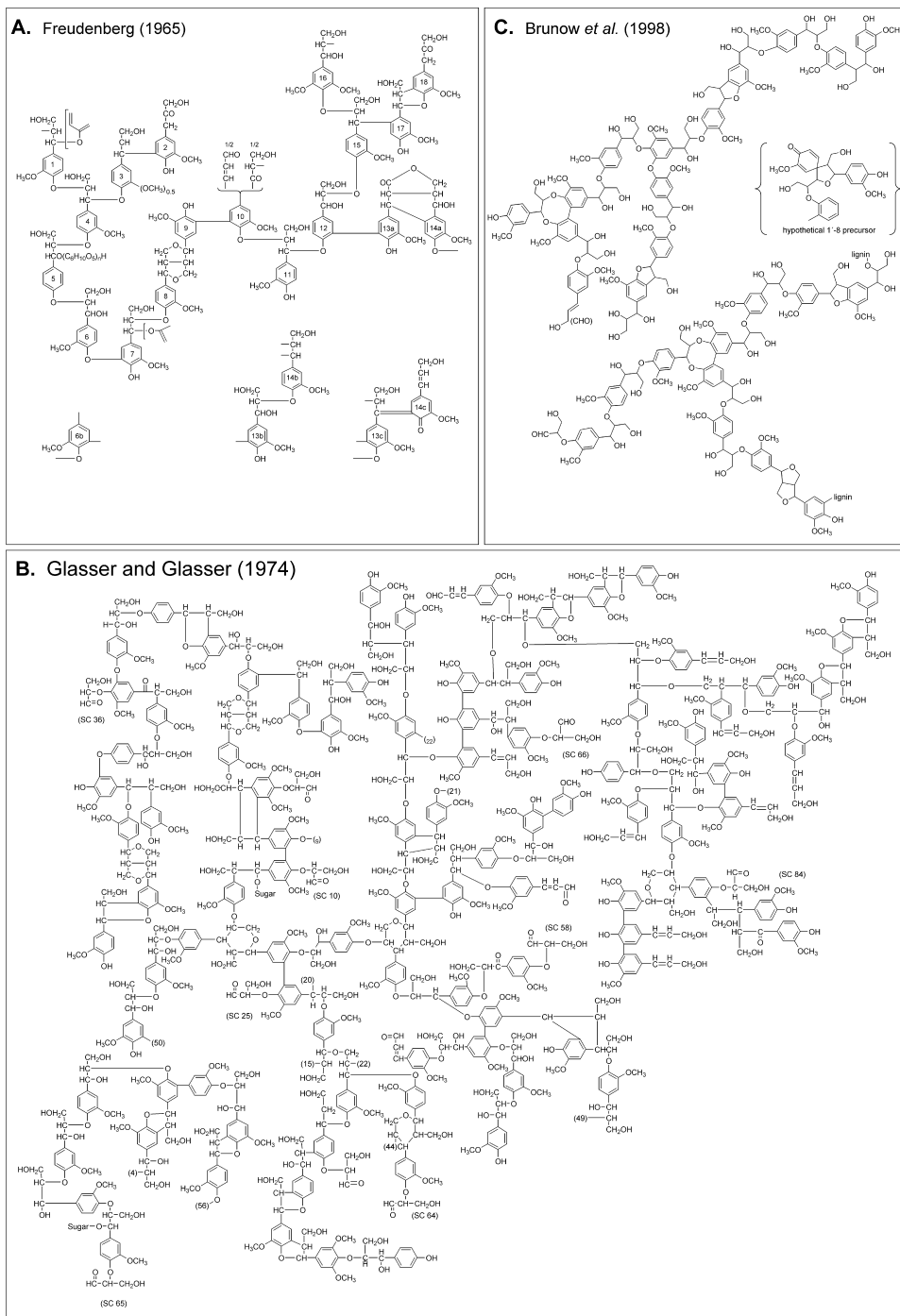


Figure 7.3 Artistic depictions of lignin structures as envisaged by (A) Freudenberg (12); (B) Glasser and Glasser's computer simulation [Redrawn from Hall *et al.* (14)] and (C) Brunow *et al.* (15).

7.2 Lignin pathway evolution, deposition, and function in vascular anatomical development

7.2.1 Vascular plant diversification and lignification

The highly diversified terrestrial (plant) environment found today originated with the shift of photosynthetic organisms from water to land some 475 million years ago (1–3). After small flora with prevascular water-conducting cells had been established in coastal wetlands, plants next developed tracheids (xylem) (22, 23), and the ability to fortify these cell walls with lignin. This evolutionary step thus gave plants a profoundly different and more efficient way to control water uptake, use and storage, thereby allowing them to become independent of wetlands and diversify onto most landscapes on earth. Further evolution of the phenylpropanoid pathway also led to the development of different types of both lignins and lignified cells, which in turn allowed plants to evolve arborescent growth in order to compete for light and space. By developing diverse height and water usage strategies, vascular plants thus created a myriad of environments that other organisms were then able to inhabit and co-evolve within. While this diversity in cell wall structure is still not fully understood, it reflects a rich source of genetic information that is central to finding methods to alter plant structure for human use without adversely affecting the subject plants (24).

The evolutionary appearance of plant vascular anatomy is quite well represented in both the fossil record and in living plants. General trends in tracheid (i.e., water-conducting cells) secondary cell wall thickenings related to the development of lignification can be seen throughout the vascular plant lineages including the extant lineages shown in Figure 7.4 (in the order their ancestors appeared in the fossil record) (2, 25–27): 1) Lycopodiophyta, represented by *Lycopodium tristachyum*, 2) Equisetophyta and Psilotophyta, with *Psilotum nudum* as an example, 3) Filicophyta, with *Pteridium aquilinum* representing higher ferns, 4) gymnosperms, e.g., loblolly pine (*Pinus taeda*), and 5) angiosperms, as represented by alfalfa (*Medicago sativa*). Tracheids with simple (limited) secondary thickenings, such as the annular and helical formations of proto- and meta-xylem are found in all these plant groups, with the reticulate thickening form found more often in the older lineages, such as the Filicophyta (arrowheads in corresponding images, right-hand column, Figure 7.4). Secondary cell wall thickenings with scalariform pitting also appeared throughout vascular plants, but with greater prevalence within the earlier lineages (through the Filicophyta); scalariform pitting has a ladder-like appearance due to a broad surface area of secondary thickening with elongated pits (arrowhead, Figure 7.4). This form is intermediate to the annular/helical forms described above and the more continuous secondary cell wall thickenings with simple (to more elaborate; not shown) circular pitting of tracheids in gymnosperms (e.g., *Pinus taeda*) and vessels of (secondary growth) angiosperms (28), such as in alfalfa (see corresponding simple pitting figure). Additionally, non-water conducting cell types with variable amounts of lignification/secondary cell walls include sclerenchyma and other structural fibers, these bearing an important mechanical function in stem and branch support. Thick-walled sclerenchyma are also found in some species of the Lycopodiophyta (e.g., the sclerified cortex of *L. tristachyum*), Psilotophyta (e.g., the sclerified outer cortex of *P. nudum*), Filicophyta (e.g., the sclerified hypodermis of *P. aquilinum*), gymnosperms, and angiosperms. Fibers with true lignin are found only in the gymnosperms and angiosperms (e.g., sclerified fibers of *M. sativa*, Figure 7.4).

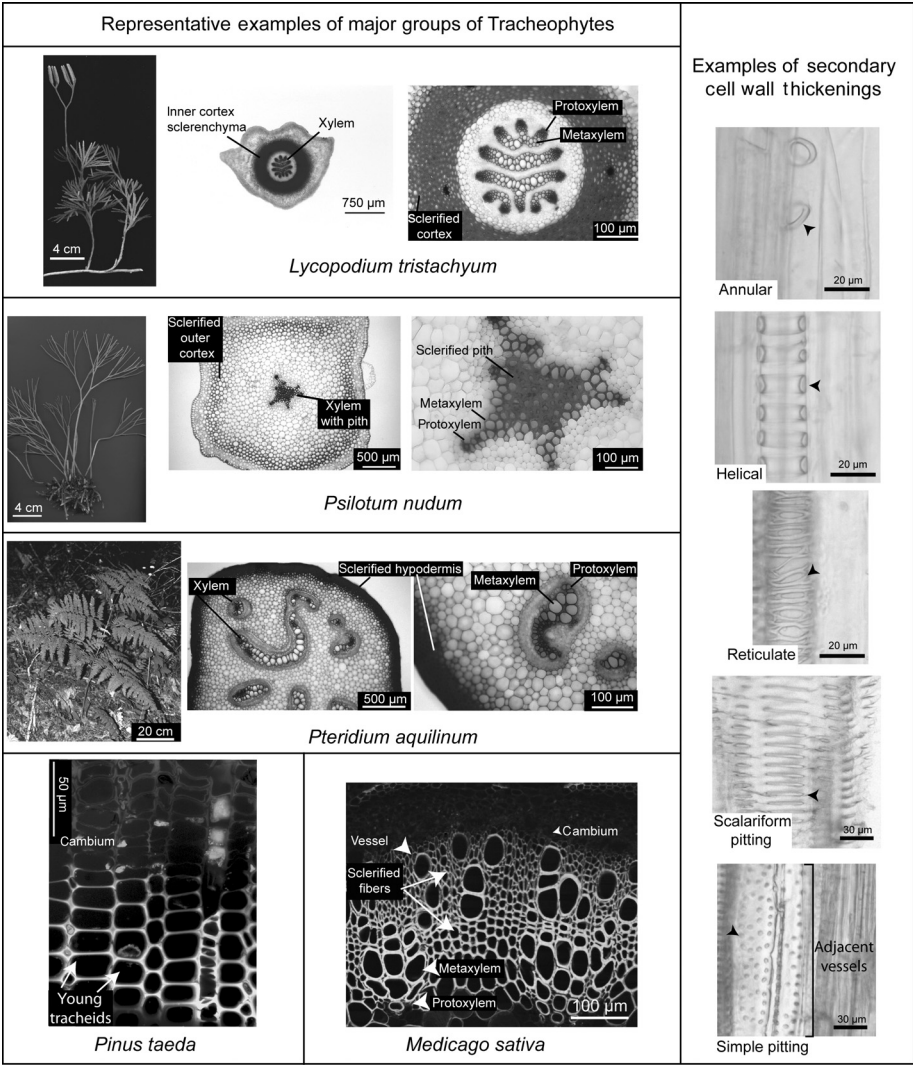


Figure 7.4 Lignified vascular and sclerenchyma anatomical structures in extant plant lineages. Proto- and metaxylem of the major plant lineages: Lycopodiophyta (e.g., *Lycopodium tristachyum*), Psilophyta (e.g., *Psilotum nudum*), Filicophyta (e.g., *Pteridium aquilinum*), gymnosperms (e.g., *Pinus taeda*), and angiosperms (e.g., *Medicago sativa*) have very similar annular and helical secondary cell wall thickenings. Some members of each group, but especially the Filicophyta, have reticulate tracheid cell wall thickenings, a more complex form than the helical form. Secondary cell wall thickenings with scalariform pitting may be found in all plant lineages and represent an intermediate form to the simple pitting found in higher plants (i.e., gymnosperms and angiosperms). Brightfield microscopy images of *L. tristachyum*, *P. nudum*, and *P. aquilinum* were taken of hand-cut fresh sections stained with phloroglucinol-HCL to reveal patterns of phenolic deposition (red) in the xylem and sclerified fiber cells. Epifluorescent confocal images of *P. taeda* and *M. sativa* were made using cryosections of fresh tissue stained with a combination of acridine orange and ethidium bromide to reveal patterns of lignification. Examples of secondary cell wall thickenings were recorded using brightfield microscopy and hand-cut longitudinal sections of fresh unstained xylem from *Equisetum telmateia* (a member of Equisetophyta for annular and helical examples), *Ophioglossum reticulatum* (a member of the Filicopsida, for the reticulate example), *P. aquilinum* (another member of the Filicopsida, for the scalariform-pitted example), and *M. sativa* (an angiosperm, for the example of vessel secondary cell wall thickenings with simple pitting). (Reproduced in color as Plate 16.)

The physiological functions of the lignins are thus quite distinct from those of either the cell wall celluloses or hemicelluloses. They are formed as biopolymeric entities *within* the cell wall, and help to reinforce the plant walls of the vasculature. In this way, they enable vascular plants to both form their water/nutrient conducting cells and also to provide a means of withstanding compressive forces acting on the overall plant body. Additionally, generation of the lignified plant cell wall matrices provides a relatively formidable physical barrier to opportunistic pathogens and to other encroaching organisms.

While the “lignin” chemistry of “primitive” plants is still yet very poorly understood, it is well established that lignin composition in higher plants is essentially only derived from the three monolignols **1**, **3**, and **5** (Figure 7.1) as indicated earlier (5, 29–31). To a much lesser extent, lignification can also involve some limited participation of related *p*-hydroxycinnamyl alcohol–monolignol esters **30–32** (Figure 7.2A), such as in grasses (5, 30, 31). This conclusion follows many detailed and exhaustive analyses of a large number of different plant species over a period encompassing 5–10 decades (32, and references therein); this established that a very strong evolutionary pressure had emerged to form lignins from these moieties and not from other non-monolignol phenolics (33). Interestingly, two of the monolignols, *p*-coumaryl (**1**) and coniferyl (**3**) alcohols, which differ only in methoxylation substitution pattern at carbon-3, are the precursors of lignins in some of the primitive (extant) plants, e.g., *Psilotum nudum*, as well as gymnosperms, e.g., loblolly pine (*Pinus taeda*) (34, 35). These moieties thus became the *p*-hydroxyphenyl (H) and guaiacyl (G) aromatic ring components of their lignins (Figure 7.2B). Various gymnosperms are, of course, widely employed as commercial sources of wood for lumber and pulp/paper production, respectively, e.g., loblolly pine, black spruce, etc.

The subsequent evolution of the angiosperms (flowering plants), by contrast, resulted in additional forms of lignified cell-wall architecture (e.g., true vessels and various fiber types), as well as the elaboration of the monolignol-forming pathway to afford the dimethoxylated monolignol, *E*-sinapyl alcohol (**5**), and ultimately the syringyl (S) aromatic ring component of the lignin biopolymers (**36**) (Figure 7.2B). Interestingly, S-moieties have also been reported to occur in some extant primitive *Selaginella* plant species (37–39), as well as in other isolated non-angiosperm families including the *Dennstaedtiaceae* (40) and *Podocarpaceae* [reviewed by Gibbs (41)]. While the evolutionary significance of such observations is not yet well understood, they may represent an example of convergent evolution.

The angiosperms are also widely used by humanity; for example, many hardwoods are utilized for lumber/pulp and paper production, whereas others (e.g., rice, corn, soybean, etc.) are food crops. In addition, three other angiosperms, thale cress (*Arabidopsis thaliana*), hybrid poplar (*Populus* sp.), and alfalfa (*Medicago sativa*) are currently of considerable scientific interest: the first as a “model” plant species, the second as a possible vehicle for generating biotechnologically modified fast-growing woody crops for fiber, biofuel applications, cellulose to ethanol, etc., and the third as a source of animal nutrition/feedstock.

Localization of lignin in xylem and fiber cells has also previously been studied in a limited number of gymnosperm and angiosperm species (36, 42–49). The earlier work indicated that lignification in gymnosperms begins in the cell wall corners and then proceeds throughout the cell wall, where *p*-coumaryl alcohol (**1**, H-unit) is differentially laid down in the cell corners/middle lamella and the coniferyl alcohol (**3**, G) residues are mainly in the secondary wall layers (42, 44, 46, 47). Syringyl moieties, by comparison, in angiosperms are deposited

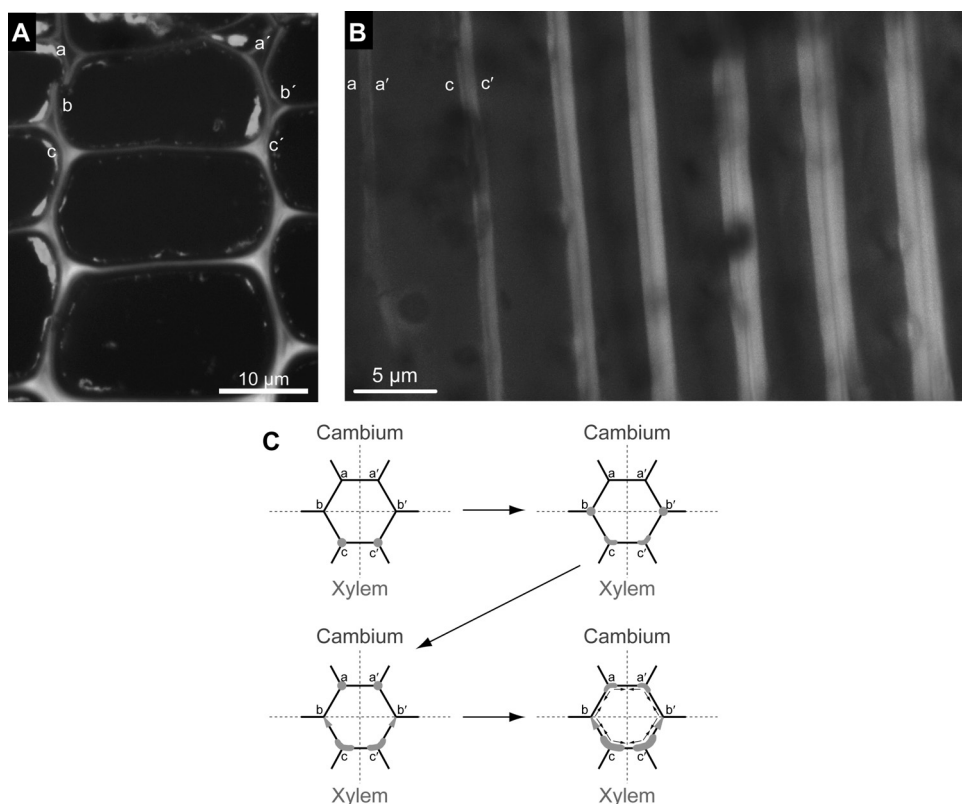


Figure 7.5 Lignification and cell wall development in loblolly pine cambial tissue (**A** and **B**) show deposition of lignin in both transverse and longitudinal sections (Y. Nakazawa, unpublished). In Fig. 5A, lignin is differentially initiated at sites a,a', b,b' and c,c'. In Fig. 5B, the particular cell has initiated lignification at cell corners c,c', but not yet at a,a'. Lignification is initiated at 3 points, c,c', b,b' then a,a' in a symmetrical but differential manner. [Cellulose/hemicellulose regions are visualized in green, and lignin is orange/yellow.] (**C**) Idealized depiction of differential lignin deposition in one tracheary cell type. → in A and • in C = lignin initiation sites. A dual-stain method using acridine orange (AO) and ethidium bromide (EB) was used to visualize lignin as a red-orange epifluorescence. The merged confocal images were individually recorded using detection filters specific to 488 nm (Emission) and 522 nm (Excitation) for AO and 568 nm (Ex.) and 598 nm (Em.) for EB. (Reproduced in color as Plate 17.)

mainly in fiber cell walls with mixed guaiacyl–syringyl lignins found in vessel cell walls (36, 45).

More specifically, the process of lignin deposition occurs in a highly organized and controlled manner, whereby the cells undergo a polarized deposition of the lignin monomers at different rates in different cell corners (as in loblolly pine developing stems, Figure 7.5A, Y. Nakazawa *et al.*, unpublished, this laboratory). Thus, as incipient xylem and sclerenchyma differentiate, *lignin initiation sites* at the cell corners/S1 sublayers of the lignifying matrix develop during maturation (see Figure 7.5A). Lignification is extended uniformly, in a continuous thread-like pattern, down the entire length of the developing tracheid toward the cambium (Figure 7.5B). This, in turn, has important biological ramifications, in terms of

both monolignol transport and monolignol (radical) alignment, and is apparently consistent with the proposed template polymerization process (discussed below) on preformed primary lignin chains (31, 50–52). Additionally, cell corners furthest from the cambial zone undergo *both lignin initiation and lignification prior to those adjacent to the cambium*, with the enlarging lignified domains in the former again being uniformly evident down the length of the tracheary element. The exact biochemical processes occurring at the lignin initiation sites is currently a subject of considerable scientific interest, as regards overall control of lignin macromolecular configuration (29).

Figure 7.5C thus shows an idealized lignification model in one cell adjacent to the cambium, whose cell corners are specified as a,a', b,b', and c,c', respectively. In this diagram, only cell corners (c,c') closest to the xylem have begun to lignify. Lignin deposition at these furthestmost cell corners then extends symmetrically along the S1 sublayer from points c and c' until the two developing zones coalesce, as well as concomitantly extending upwards to the next two adjacent cell corners (i.e., b,b' in this case). Lignin deposition is *subsequently* initiated at points b,b' and symmetrical deposition occurs again in a likewise manner. Finally, lignin deposition in the remaining two corners (a,a') closest to the cambium is initiated at the last phase, and these zones also begin to expand uniformly – albeit in a delayed manner – until eventually coalescence of lignin within the entire wall is achieved. [Figures 7.5A and 7.5B depict the asymmetry in cell wall thickness/lignin deposition as cell wall development continues, i.e., wall c,c' is thicker and more heavily lignified than a,a' at this stage.] Additionally, when cell wall development has occurred, the adjacent unlignified cell (closest to the cambium) is next “conscripted” to undergo lignin assembly/cell wall thickening as before. Such observations thus do not appear to be in agreement with Freudenberg’s original hypothesis of (random) diffusion of monolignol (glucoside) precursors into the cell walls undergoing lignification (53, 54).

In addition to their structural roles in plant stems, lignins provide a physical barrier to opportunistic pathogens (31), and various specialized structures throughout the plant body (e.g., trichomes, harboring an arsenal of plant defense compounds) apparently contain a lignified base. To put this into a more holistic perspective, Figure 7.6 illustrates some of the various tissues and cell wall types that are considered to contain lignified elements in *Arabidopsis*; these can be readily visualized through expression of the GUS-cinnamyl alcohol dehydrogenase (AtCAD4 and 5) promoter fusion product in the vascular apparatus (55), with CAD encoding the final step in monolignol biosynthesis (31). In this regard, it should also be noted that in this species these two CAD genes (AtCAD4 and 5) are considered to be largely responsible for the penultimate step(s) leading to lignification (discussed later) (56, 57). The main point is that any adverse effect on stem lignification could thus also potentially disrupt other physiological processes in these different tissues and organisms, i.e., whether in terms of structural support and/or in defense system impairment.

Interestingly, for almost three quarters of a century, various lignin mutants (beginning with the brown-midrib mutants in maize) have been described (58–67). Essentially, none currently find application as commercial cultivars, because of the deleterious effects, for example, on the overall plant vasculature, the reproductive system, and so forth. However, recent studies have begun to shed important light on the various genes and enzymes involved in each mutation, and how they affect lignin deposition processes proper. With the recent advances in lignin pathway modulation, biomechanics approaches are also beginning to be increasingly applied in order to begin to correlate such modulations with that of alterations

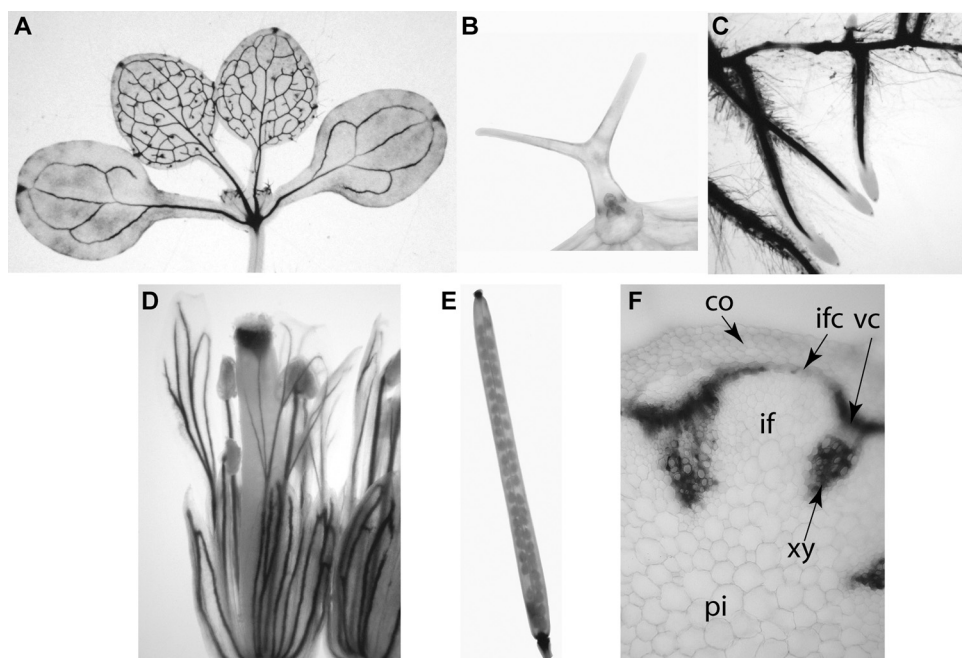


Figure 7.6 Cinnamyl alcohol dehydrogenase gene expression in the vascular apparatus of *Arabidopsis thaliana* (55). Selected examples of GUS-visualized expression of AtCAD5 (**A**, **C–F**) and AtCAD4 (**B**). (**A**) In vascular apparatus, including hydathodes of 2-week-old leaf tissue; (**B**) at the base of trichomes; (**C**) in primary and secondary roots; (**D**) in sepal/petal veins, style, anthers, and stamen filaments of the flower; (**E**) in the abscission and style regions of the silique; and (**F**) in vascular cambium, interfascicular cambium, and the developing xylem of the inflorescence stem. (Reprinted from *Phytochemistry*, vol. 68, Kim, S.-J., Kim, K.-W., Cho, M.-H., Franceschi, V.R., Davin, L.B. & Lewis, N.G., Expression of cinnamyl alcohol dehydrogenases and their putative homologues during *Arabidopsis thaliana* growth and development: Lessons for database annotations? pp. 1957–1974, Copyright 2007, with permission from Elsevier.) (Reproduced in color as Plate 18.)

of plant structural integrity (68–72). As more comprehensively discussed below, this is an essential first step toward understanding to what extent lignin composition and content can be manipulated.

7.2.2 Heartwood and reaction (compression/tension) wood tissues

Lignification and cell wall assembly processes are also frequently altered/modified by plants in response to structural and environmental stresses/conditions. For instance, many arborescent gymnosperms and angiosperms form heartwood beginning at the center of the stem and this eventually extensively radiates outwards through (in part) secretion of defense molecules (73), such as lignans from adjacent parenchyma cells in the sapwood (74) (Figures 7.7A–C). That is, *following* lignification, the various heartwood deposition processes result in formation of large quantities of diverse metabolic products, including species-specific resins and lignans, which increase the density (and presumably support) of the wood, as well as the

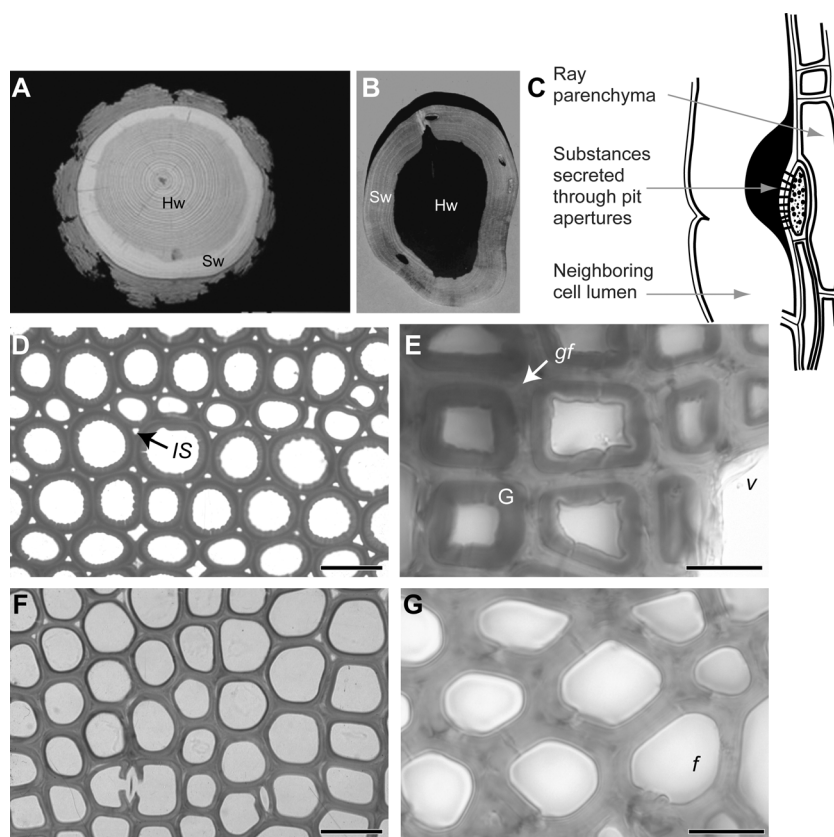


Figure 7.7 Woody cross sections showing heartwood deposition and reaction wood tissues. Heartwood in (A) tamarack (*Larix laricina*) and (B) ebony (*Diospyros ebenum*), as well as (C) secretion of heartwood constituents by ray parenchyma cells into lumen of neighboring cells; this appears to occur through pit apertures (73). Light micrograph cross section of compression (D) and “normal” (F) wood in Douglas fir (*Pseudotsuga menziesii*) (75) and of tension (E) and “normal” (G) wood in black cottonwood (*Populus balsamifera* ssp. *trichocarpa*) (72). Bar: 20 μm (D, F) and 10 μm (E, G). Abbreviations: f, fiber; G, G-layer; Hw, heartwood; is, intercellular space; gf, gelatinous fiber; Sw, sapwood; v, vessel. [Reprinted from (A) *Phytochemistry*, vol. 57, Kwon, M., Bedgar, D.L., Piastuch, W., Davin, L.B. & Lewis, N.G., Induced compression wood formation in Douglas fir (*Pseudotsuga menziesii*) in microgravity, pp. 847–857, Copyright 2001, with permission from Elsevier. (B) *Current Opinion in Plant Biology*, vol. 2, Lewis, N.G., A 20th century roller coaster ride: A short account of lignification, pp. 153–162, Copyright 1999, with permission from Elsevier. (C) *ACS Symposium Series*, vol. 697, Gang, D.R., Fujita, M., Davin, L.B. & Lewis, N.G., The “abnormal lignins”: Mapping heartwood formation through the lignan biosynthetic pathway, pp. 389–421, Copyright 1998, with permission from American Chemical Society. (E and F) *The American Journal of Botany*, vol. 94, Patten, A.M., Jourdes, M., Brown, E.E., Laborie, M.-P., Davin, L.B. & Lewis, N.G., Reaction tissue formation and stem tensile modulus properties in wild type and *p*-coumarate-3-hydroxylase downregulated lines of alfalfa, *Medicago sativa* (Fabaceae), pp. 912–925, Copyright 2007, with permission from the Botanical Society of America.] (Reproduced in color as Plate 19.)

resistance of such tissues to biodegradation (lignocellulose deconstruction) (74). Such substances thus often help confer additional defensive properties to heartwood, making them more resistant to pathogen challenges and helping to increase longevity. Often particular types of defense metabolites (such as lignans) are found in specific tree species, further demonstrating the quite remarkable chemical diversity that has evolved through the phenylpropanoid pathway, e.g., the lignan – and other metabolite – enrichment of western red cedar (*Thuja plicata*) contributes to its life span that can potentially exceed 3000 years or so. In other species, such as poplar, a lignan-enriched tissue is also formed, which is sometimes referred to as a “ripewood.”

Lignification/cell wall formation processes are also continuously modulated when woody plant species form branches, are fast-growing, and/or are challenged by having their stems bent (as when growing on a slope). The so-called “reaction wood” formed is trivially known as compression wood (Figure 7.7D) in gymnosperms and tension wood (Figure 7.7E) in angiosperms; Figures 7.7F and 7.7G show “normal” wood for comparison purposes. Both types of reaction wood also show variable changes in cellulose content but, interestingly, the H-lignin contents of compression wood are increased while that of overall lignin contents in tension wood are decreased in some species. Furthermore, the means for achieving both stem and branch orientation differ profoundly between gymnosperms and angiosperms; the former produce the lignin-rich reaction (compression) wood on the underside of stems/branches (75, 76), whereas the latter form the reaction (tension) wood with variable levels of lignin deficiency on the upper-sides of both (30, 72) (discussed below). One purpose of these (re-)orientation mechanisms appears to be to enable maximization of the exposure of the photosynthetic apparatus within the leaves for efficient photosynthetic capture of energy from the sun. The underlying biochemical/molecular mechanisms and reasons for such different lignin/cell wall forming responses to similar mechanical challenges are, however, not yet known. Nevertheless, these examples illustrate some of the quite remarkable changes that plants can undergo in order to obtain either enhanced protection of the lignocellulosic matrix and/or in modifying growth/development through programmed modulation of plant cell wall assemblies.

7.3 Pioneers of monolignol biosynthesis, recent progress, and metabolic flux analyses

Both the monolignol and the shikimate–chorismate biochemical pathways have been extensively studied over a period spanning nearly five decades. As regards the former, the reader is highly encouraged to review the two most comprehensive treatises on monolignol biosynthesis (31) and monolignol pathway genetic manipulation (77). The first provides an historical account of the pioneering work in enzyme identification, enzyme isolation, and subsequent gene cloning in the monolignol/lignin pathway, whereas the second largely attempts to identify predictable trends in downregulating/mutating various pathways steps, e.g., following application of standard molecular biological approaches. Interestingly, the five-decade held view that lignins were randomly assembled had apparently all but deterred many researchers in this field from systematically studying – in a productive and predictive hypothesis-driven manner – the effects of manipulating the lignin-forming apparatus.

Moreover, although there are occasional reports describing the monolignol pathway as being recently redrawn in the past few years (78), in hindsight this is really not the case. Instead, a number of discoveries had been made – some up to a quarter of a century ago – whose significance had not yet been appreciated by various researchers at the time. We, therefore, briefly review the monolignol pathway, the highlights of recent years, and the recent advances made in understanding the nature of metabolic pathway flux associated with monolignol/lignin formation. Furthermore, because of the completion of the *Arabidopsis* genome sequencing in 2000 (79), an emphasis is also placed on the study of the phenylpropanoid-forming biochemical machinery in that organism as it is currently the most extensively studied.

7.3.1 Phenylalanine formation

There has long been uncertainty as to how plants produce the (essential) aromatic amino acid, phenylalanine (6) (Figure 7.8). Originally considered to be formed via transamination of phenylpyruvate (35) (80), work pioneered by Jensen and colleagues (81) and later by Siehl and Conn (82) provided evidence for an alternate pathway using arogenate (37); however,

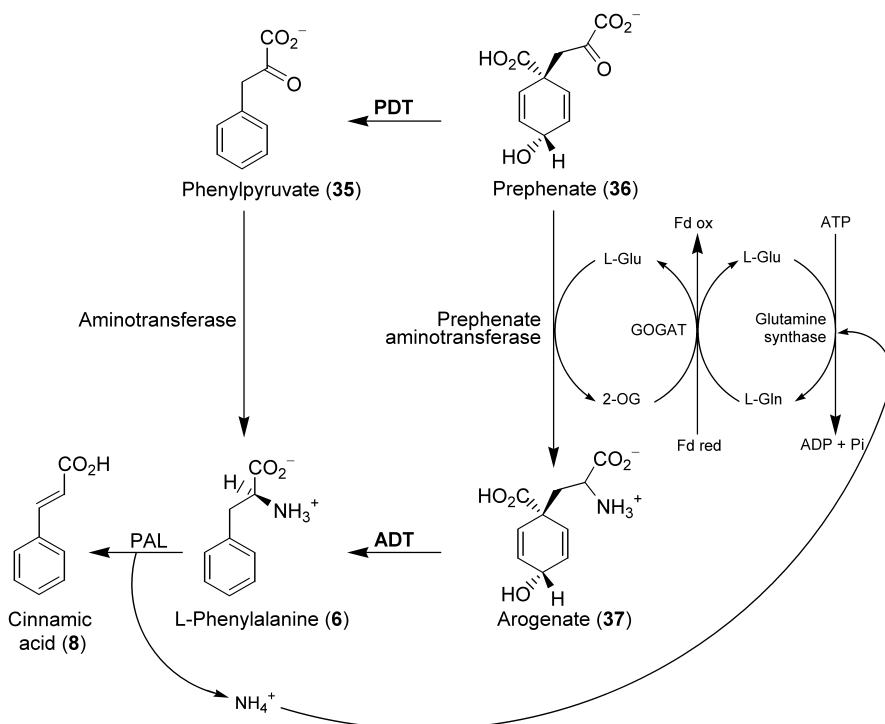


Figure 7.8 Proposed biosynthetic pathway from prephenate (36) and arogenate (37) to Phe (6) in plants (83) and nitrogen recycling mechanism (85–88).

the aroenate dehydratases detected in crude extracts were apparently not stable enough for purification and there were no encoding genes identified.

In *Arabidopsis*, using a data mining approach, we identified six putative aroenate dehydratases (ADTs) and/or prephenate dehydratases (PDTs) based on (relatively low) sequence homology to known bacterial PDTs (83). Subsequent expression of each of the recombinant proteins in a NUS-His-tagged form established that each was capable of more efficiently converting aroenate (37) into phenylalanine (6), rather than catalyzing the conversion of prephenate (36) into phenylpyruvate (35) (83), i.e., indicative of a six-membered aroenate dehydratase family in *Arabidopsis*. It will be instructive to next establish the nature of this Phe-forming metabolic network. That is, by determining which of these genes are involved in phenylpropanoid/lignin metabolism, protein formation, and/or other biochemical processes, as well as to what extent this gene family is functionally redundant (due to co-expression).

7.3.2 Metabolic flux analyses and transcriptional profiling in the monolignol pathway

These analyses have been most useful in predicting the outcome of various manipulations in the lignin-forming pathway. That is, previous studies, whereby monolignol **1** and **3** formation could be induced in loblolly pine (*Pinus taeda*) cell suspension cultures, enabled us to gain important insights into factors controlling metabolic flux to both *p*-coumaryl (**1**) and coniferyl (**3**) alcohols (34, 35). Thus, by increasing levels of available sucrose, the monolignol-forming pathway could be induced, with the cells secreting the monolignols (in the presence of an H₂O₂-scavenger) into the culture medium.

The data (based on both measuring various pathway metabolite levels and transcript profiles) provided quite informative insights: the first was that (regulation of) carbon allocation to the pathway was controlled upstream through the amounts of Phe (6) being made available, as well as through the differential activities of both cinnamate-4-hydroxylase (C4H) and *p*-coumarate-3-hydroxylase (*p*C3H). Furthermore, metabolite analyses also indicated that formation of both *p*-coumaryl (**1**) and coniferyl (**3**) alcohols could be differentially induced, suggesting the existence of distinct metabolic control over segments (i.e., H versus G) within the monolignol/lignin forming processes through differential modulation of *p*C3H activity. Beyond the hydroxylation steps, other downstream enzymatic steps (see Figure 7.1) were not considered to be rate limiting, at least under the conditions employed in the studies. [Of course, any enzymatic step becomes rate limiting if abolished or “knocked out.”] Transcriptional profiling data of each of the known steps involved in monolignol biosynthesis (available at that time) also appeared to support this analysis and interpretation (35).

7.3.3 Phenylalanine and tyrosine ammonia lyases

Phenylalanine ammonia lyase (PAL) represents the first committed step to phenylpropanoid metabolism, and thus to a wide range of phenolic products (lignins, lignans, hydroxycinnamic acids, flavonoids, suberins, etc.). Discovered by Koukol and Conn (84) in 1961, this enzymatic conversion required no *added* cofactor. Moreover, the ammonium ion released

during the deamination step had also been subsequently demonstrated to be recycled via GS/GOGAT, with the glutamate so formed then serving as the amino donor for aroenate (35) formation (Figure 7.8) (85–88). In this way, the means to both recycle the nitrogen and sustain phenylpropanoid metabolism can occur without any apparent further need for an additional nitrogen source. A gene encoding a PAL was reported by Edwards *et al.* in 1985 (89); an earlier description by others (90, 91) of PAL cloning was later shown not to be PAL (89, 92).

In *Arabidopsis*, there are four *bona fide* PAL gene family members present [as determined by both analysis of the genome sequence (93), and also by recombinant protein characterization (94)]. In terms of its biochemical mechanism, the X-ray crystal structure of PAL was also recently described (95); additionally, it was deduced that a highly conserved tripeptide sequence (Ala-Ser-Gly) in PAL undergoes spontaneous dehydration/cyclization to give MIO (3,5 dihydro-5-methylidene-4*H*-imidazol-4-one) (95) based on earlier studies with histidine ammonia lyase (HAL) (96, 97). MIO is envisaged to serve as an internal cofactor for catalysis. Interestingly, other studies using both tyrosine ammonia lyase (TAL) (from *Rhodobacter sphaeroides*) (98) and PAL (from *Arabidopsis*) (99) have demonstrated that the interconversion of PAL to TAL activity can be effectuated by a single amino acid substitution (i.e., H89F converts TAL to PAL in *R. sphaeroides*, whereas in *Arabidopsis*, the PAL/TAL switch is F144H).

7.3.4 Cytochrome P-450s and hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyl transferases

As indicated in Figure 7.1, the phenylpropanoid pathway has several hydroxylation steps leading to the monolignols 1–5. In this context, all three enzymatic steps, in terms of historical discovery and recent progress made, were comprehensively discussed in Lewis *et al.* (31) and Anterola and Lewis (77). The first two of these, C4H and *p*C3H, are common to all known vascular plant species, whereas the third “ferulate-5-hydroxylase” (F5H) is restricted mainly to the angiosperms. An apparent exception to this is that some members of the “primitive” *Selaginella* family also form syringyl lignins (37–39), this being attributed to an F5H homologue (100).

All three conversions are now established to be cytochrome P-450 catalyzed (with each utilizing an associated NADPH reductase). Of these, C4H was discovered by Russell and Conn in 1967 (101), with the encoding gene reported in 1993 by Teutsch *et al.* (102) and Mizutani *et al.* (103). *p*C3H, on the other hand, was the subject of much literature confusion, since two other enzymatic processes were reported as being responsible for caffeic acid (10) formation: namely, a polyphenol oxidase (104, 105) and a FAD-dependent oxidase (106). These conversions have never been established though to have any physiological role as regards biosynthesis of caffeate derivatives leading to monolignol 2–5/lignin formation. On the other hand, the seminal contributions made by Heller and Kühnl in 1985 (107) had established that, using parsley (*Petroselinum crispum*) cell suspension cultures, *p*-coumaroyl shikimate (25) was converted into the corresponding caffeoyl shikimate derivative (27) through the action of a cytochrome P-450. [By contrast, *p*-coumaric acid (9) was not converted into caffeic acid (10).] Yet, it was not until much later that the true significance of this discovery was fully recognized [see Lewis *et al.* (31)]. That is, they had, in fact, identified

both the enzymatic nature of the hydroxylase, as well as the substrate(s) being utilized, as early as 1985. It was not until 2001, however, that a cytochrome P450 gene encoding *pC3H* was reported by Schoch *et al.* (108). In agreement with Heller and Kühnl (107), the heterologously expressed protein was able to convert *p*-coumaroyl shikimate (25) into the corresponding caffeoyl (27) derivative, but not *p*-coumaric acid (9) into caffeic acid (10).

Since both C4H and *pC3H* are endoplasmic-reticulum (ER) membrane bound, it is quite interesting that the initial hydroxylation product, *p*-coumaric acid (9), needs first to be esterified prior to hydroxylation at the carbon 3 position. What that specifically means as to how the H versus G/S pathway segments are spatially and temporally kept distinct though still awaits further clarification. Yet, as for the *pC3H* discovery of Heller and Kühnl (1985) (107), the formation of *p*-hydroxycinnamate shikimate esters had also long been known, albeit overlooked. These were considered to be possible phenylpropanoid pathway intermediates as early as in the 1970s, i.e., but again without initial recognition by many other researchers. Indeed, hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase (HCT) was first detected by Stöckigt and Zenk (1974) (109). Only much later (110, 111) (Cardenas *et al.*, this laboratory, unpublished results), were the genes encoding both HCT, and the closely related hydroxycinnamoyl CoA:quinic acid hydroxycinnamoyl transferase (HQT), cloned from tobacco, and their roles in monolignol metabolism defined. Interestingly, HCT preferentially utilizes shikimate (29) as an acyl group acceptor, whereas HQT prefers the quinate (28) moiety instead.

The final cytochrome P-450 hydroxylation step, F5H, was initially discovered by Grand in 1984 (112) and was considered to utilize ferulic acid (11) as substrate to afford 5-hydroxyferulic acid (12); the latter had previously been discovered in *Zea mays* and *Hordeum vulgare* (113). Later, the group of Fukushima *et al.* (114, 115) demonstrated that when [9-²H₂,OC²H₃] coniferyl alcohol (3) was administered to *Magnolia kobus* root tissue, it was metabolized into pentadeuterated-syringyl units with two 9-deuterium atoms via hydroxylation at the C-5 position; moreover, the deuterium label was also present in the S-lignin that was formed. These seminal findings were subsequently confirmed by re-examination of the catalytic properties of a recombinant *Arabidopsis* F5H (116) where it was established that both coniferyl alcohol (3) and coniferyl aldehyde (21) served as preferred substrates over ferulic acid (11). In apparent agreement with possibly defining roles in carbon allocation in the pathway, it is of interest that in *Arabidopsis* there are only 1 C4H, 1 *pC3H*, and 1 HCT genes, as well as possibly 2 F5Hs, i.e., in contrast to the 6- and 4-multigene ADT and PAL families.

7.3.5 4-Coumarate CoA ligases

Gross and Zenk in 1966 (117) provided the first experimental evidence that metabolism of hydroxycinnamic acids 9, 11, and 13 occurred through the corresponding CoA esters, this then being confirmed in other studies with AMP derivatives being reported as transient intermediates (118–120). It was not, however, until 1987 when the first gene encoding a *bona fide* 4CL was described (121). Unlike the preceding enzymatic steps that displayed a very narrow range of substrate versatility, all known 4CLs showed fairly broad substrate specificities [i.e., for *p*-coumaric (9), caffeic (10), ferulic (11), and 5-hydroxyferulic (12) acids], but not for sinapic acid (13). For example, in *Arabidopsis*, although some 14 genes had previously been annotated as 4CLs and/or putative 4CLs, only four of these actually

had 4CL catalytic activity, with each displaying broad albeit distinctive substrate versatility (122). In particular, although there has long been speculation that a sinapic acid (**13**) specific 4CL might be operative for S-lignin formation, no experimental evidence for this could be obtained – at least in *Arabidopsis* (122).

7.3.6 Cinnamoyl CoA reductases and cinnamyl alcohol dehydrogenases

Both cinnamoyl CoA reductases (CCRs) and cinnamyl alcohol dehydrogenases (CADs) are substrate versatile NADPH-dependent enzymes. These proteins were detected and purified in the laboratories of Zenk and coworkers (118, 120, 123) in the early 1970s; CCR is a type-B reductase abstracting the 4-pro-*S*-hydrogen of NADPH (124), whereas CAD is a type-A reductase (4-pro-*R*-hydrogen abstraction) (123). As for 4CLs, all known CCRs and CADs are quite substrate versatile: CCR can catalyze the in vitro reduction of *p*-coumaroyl (**14**), caffeoyl (**15**), feruloyl (**16**), 5-hydroxyferuloyl (**17**), and sinapoyl (**18**) CoAs into the resulting aldehydes **19–23**, and the first gene encoding a CCR was described in 1997 (125). CADs are able to catalyze the conversion of aldehydes **19–23** into the monolignols **1–5**, with the first *bona fide* gene encoding a CAD reported by Knight *et al.* (126); earlier reports of CAD being cloned (127) later turned out to be malic enzyme (128, 129).

In the *Arabidopsis* Information Resource (TAIR), there were some 11 genes annotated as CCR/CCR-like, and another 17 as CAD/CAD-like (93). To date, only two CCRs have been demonstrated to have both this activity and physiological function (130–132). A similar situation exists for CADs: two of these (AtCAD4 and AtCAD5) are the most catalytically active, and another four (AtCAD2, 3, 7, and 8) have very low overall activities making their involvement in lignification/monolignol formation suspect (56); additionally, AtCAD1, 6, and 9 did not have any detectable CAD activities proper in vitro, and thus their biochemical function currently remains unknown (56).

The biochemical mechanism of CAD has also been the subject of considerable interest, and the X-ray crystal structure of AtCAD5 was recently described (133). Site-directed mutagenesis has implicated Glu70 in catalysis, the latter being coordinated with the catalytic Zn²⁺ through its side chain (Figure 7.9A). A proton shuttle mechanism for CAD has also been provisionally proposed (Figure 7.9B) but remains to be experimentally proven.

7.3.7 COMTs and CCOMTs

As shown in Figure 7.1, the pathway to the monolignols **3–5** includes two methylation steps. These have been shown to be catalyzed by caffeic acid *O*-methyltransferases (COMTs) and caffeoyl CoA *O*-methyltransferases (CCOMTs). COMT activity was first detected in cambial tissue isolated from an apple tree species by Finkle and Nelson (134). It was thought for much time that COMTs could catalyze the formation of both ferulic (**11**) and sinapic (**13**) acids from caffeic (**10**) and 5-hydroxyferulic (**12**) acids, respectively. However, in 1989, Pakusch *et al.* (135) described an enzyme, CCOMT, capable of converting caffeoyl CoA (**15**) into feruloyl CoA (**16**). Additionally, downregulation of the COMT gene later unambiguously

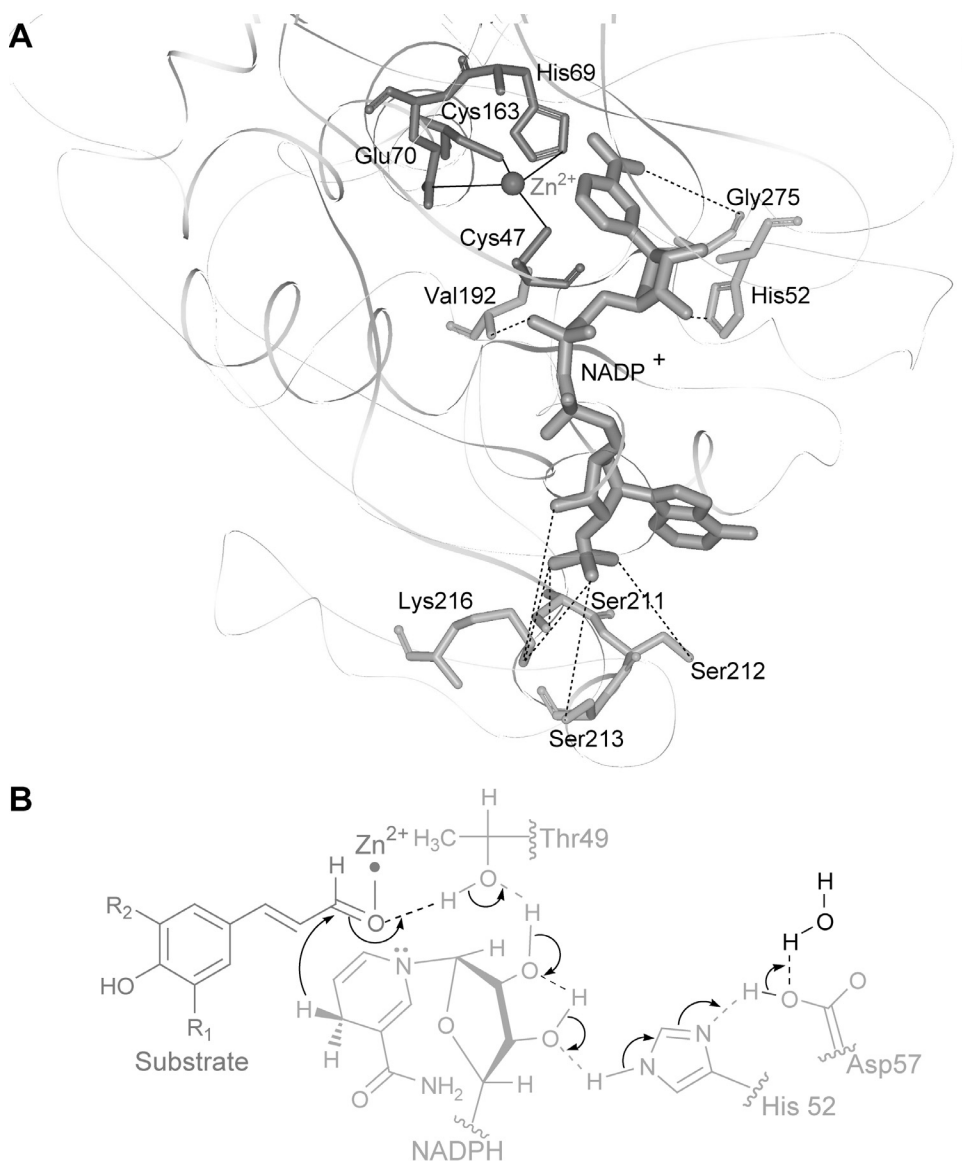


Figure 7.9 (A) Structure of the substrate-binding pocket of NADP⁺ binary form of AtCAD5 showing the catalytic Zn²⁺ ion (red sphere) tetrahedrally coordinated by Cys47, His69, Cys163 and Glu70 (blue). The NADP⁺ molecule (orange) is held by Val192, Ser211, Ser212, Ser213, Lys216 and Gly275 (green) (133). [Possible hydrogen bonds are shown as black dotted lines.] (B) Proposed proton shuttling mechanism during the reduction process in the active site of the AtCAD5. Solid arrows indicate the movement of two electrons among the functional groups during substrate reduction (133). The possible hydrogen bonds involved are shown with dotted lines. (Reprinted from Organic and Biomolecular Chemistry, vol. 4, Youn, B., Camacho, R., Moinuddin, S.G.A., Lee, C., Davin, L.B., Lewis, N.G. & Kang, C., Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4, pp. 1687–1697, Copyright 2006, with permission from The Royal Society of Chemistry.) (Reproduced in color as Plate 20.)

established its role in syringyl (S) unit formation and not the initial step involving G-lignin deposition (136). Genes encoding COMT and CCOMT were first reported in 1991 by Bugos *et al.* (137)/Gowri *et al.* (138) and by Schmitt *et al.* (139), respectively, with the actual biochemical/biophysical function for COMT being determined later by Atanassova *et al.* (136).

In the TAIR database, there are 17 genes putatively annotated as COMTs and five as CCOMTs. Investigations in our laboratory have indicated that out of the 17 COMTs only one, AtCOMT1, was capable of methylating caffeic (**10**)/5-OH ferulic (**12**) acids, caffeyl (**20**)/5-hydroxyconiferyl (**22**) aldehydes and caffeyl (**2**)/5-hydroxyconiferyl (**4**) alcohols *in vitro*. The efficiency of the reaction is higher when the aldehydes **20** and **22** are used as substrates (Zhang *et al.*, manuscript in preparation). Only two CCOMTs, AtCCOMT1 and AtCCOMT2, have been investigated: caffeoyl (**15**) and 5-hydroxyferuloyl (**17**) CoAs are the preferred substrates for AtCCOMT1, with **10** and **12** not being converted into ferulic (**11**) and sinapic (**13**) acids, respectively. AtCCOMT2 more efficiently methylated quercetin (Takahashi *et al.*, manuscript in preparation).

7.3.8 Proteins of unknown physiological/biochemical functions in monolignol metabolism, “CAD1” and “sinapyl alcohol dehydrogenase, SAD”

7.3.8.1 CAD1

In addition to the *bona fide* CADs [such as from tobacco (126), loblolly pine (35) and *Arabidopsis* (56, 57)], there continue to be a number of reports proposing that another class of dehydrogenases (so-called CAD1) is involved in monolignol/lignin biosynthesis (140, 141). These enzymes, by contrast, bear little homology to *bona fide* CADs and lack both catalytic and structural zinc metal ions (56). For example, a dehydrogenase, annotated as CAD1, was purified from *Eucalyptus gunnii* in parallel with a *bona fide* CAD, then called CAD2 (142). CAD1 was shown to have lower affinity for *p*-coumaryl (**19**)/coniferyl (**21**) aldehydes ($K_m = 70$ and $25 \mu\text{M}$) as compared to CAD2 ($K_m = 1.2$ and $1.7 \mu\text{M}$), with sinapyl aldehyde (**23**) not serving as a substrate; CAD1, however, was not characterized further. It was also presumed to be a monomer (~ 38 kDa), as compared to CAD2, which is a dimer (~ 83 kDa) (142). However, the encoding gene (*CAD1-5*) (140) had only 30/15% similarity/identity to *bona fide* CADs [e.g., AtCAD5 (56)].

Since then, two other alcohol dehydrogenase genes (so-called *NtCAD1-1* and *NtCAD1-7*) present in tobacco were also obtained by screening a cDNA library with the *E. gunnii* *CAD1-5* coding sequence. Curiously, these were reported as a new step for the biochemical formation of coniferyl alcohol (**3**) (141). Both dehydrogenases again lacked the Zn catalytic center and the Zn-binding signature found in *bona fide* CADs, and also had $\sim 30\%$ similarity and $\sim 15\%$ identity compared to AtCAD5. The kinetic parameters of both *NtCAD1-1* and *NtCAD1-7* were nevertheless investigated using coniferyl aldehyde (**21**)/alcohol (**3**) and sinapyl alcohol (**5**) as potential substrates. Both proteins were apparently capable of catalyzing the forward reaction of reduction of coniferyl aldehyde (**21**), as well as the oxidation of coniferyl alcohol (**3**), but they were apparently unable to convert sinapyl alcohol (**5**) to sinapyl aldehyde (**23**); the forward reaction was not examined. The K_{enz} values for both

NtCAD1-1 and *NtCAD1-7* were very modest (low) relative to *bona fide* CADs, i.e., 3230 and 7650 M⁻¹ s⁻¹ versus 348 000 M⁻¹ s⁻¹ for *AtCAD5*. We would suggest that this family of dehydrogenases, even though they may be able to inefficiently reduce coniferyl aldehyde (**21**) to coniferyl alcohol (**3**) in vitro, have biochemical functions unrelated to that of monolignol/lignin biogenesis. Indeed, as discussed below, a recent study has also reported that genes highly homologous to “CAD1” are instead 2-phenylacetaldehyde reductases. As a further caution, numerous dehydrogenases are known to exhibit broad substrate versatility of varying levels of efficacy.

7.3.8.2 Sinapyl alcohol dehydrogenase

Another distinct dehydrogenase, this time from aspen, was claimed to be specific for sinapyl alcohol (**5**)/syringyl lignin formation (143). This report from the Chiang laboratory was quite unexpected. This was because its actual broad substrate versatility for cinnamyl aldehydes **19–23** eliminated it as being biochemically-specific for sinapyl alcohol (**5**)/syringyl lignin formation (77). That is, and as previously noted for all the CADs proper, this enzyme was also substrate versatile. Thus, any substrate specificity, if it exists at all, would presumably be a result of compartmentalization, i.e., where sinapyl aldehyde (**23**) and the so-called “sinapyl alcohol dehydrogenase” were co-localized. Since this has not been established, the physiological role of this dehydrogenase remains unknown at present. More recently, an X-ray crystal structure for “SAD” was also obtained (144), but which contained a vastly different substrate-binding pocket (in terms of both size and amino acid composition) to that of a *bona fide* CAD, i.e., *AtCAD4/5* (133). Specifically, only 2 of the 12 amino acid residues which constitute the CAD substrate-binding pocket were conserved in “SAD,” with the substrate-binding pocket for the latter also being considerably larger (133). Taken together, these data suggest an alternative biochemical function for “SAD,” as also provisionally suggested by Bomati and Noel (144).

Furthermore, in *Arabidopsis*, no evidence for any requirement for a “SAD” was obtained, since $\geq 94\%$ of all monolignol **1–5** formation for lignification was carried out by *AtCAD4/5*, both of which share considerable homology (74–83% similarity) to that of *bona fide* CADs (56). Moreover, the dehydrogenases of highest similarity to the “sinapyl alcohol dehydrogenases” (i.e., *At4g37970* and *At4g37980*) did not reduce the *p*-hydroxycinnamaldehyde **19–23** to afford the monolignols **1–5** to any considerable extent in vitro (56). Thus, as for the other dehydrogenases described above, clarification of their biochemical and physiological roles need to be established as well.

A rice mutant, *gold hull and internode* (*gh*), was also first described as early as 1917, and characterized by a reddish-brown color in the hull and internode but not in the midrib. Since then a series of mutants were identified, i.e., *gh1–gh4*. Zhang *et al.* (145) recently characterized the *gh2* mutant and showed that the *GH2* gene encodes a CAD. The substrate versatile kinetic properties of the recombinant GH2 proteins were determined showing K_{enz} values for coniferyl (**21**) and sinapyl (**23**) aldehydes of $\sim 289\,000$ versus $\sim 162\,000$ M⁻¹ s⁻¹, respectively, these being in relatively good agreement with kinetic values for *AtCAD5* in *Arabidopsis* (56). The recombinant mutant *gh2* protein, however, lost the corresponding activities. Additionally, analyses of CAD activity in all tissues (i.e., panicle, hull, blade, midrib, sheath, internode, and root) of both wild type and *gh2* mutant plants showed that

the formation of coniferyl alcohol (3) was greatly reduced in the roots, internodes, hulls, and panicles of the *gh2* mutant, with no detectable formation of sinapyl alcohol (5). On the other hand, the *gh2* mutant had apparently little to no effect on overall lignification (estimated at a 5–6% reduction). Taken together there is no evidence thus far that “SAD” has the specific biochemical/physiological functions reported earlier (143).

7.4 Recent developments: metabolic networks in the monolignol/lignin forming pathway (*Arabidopsis*) and (current) database annotations/limitations – opportunities and challenges

The monolignol pathway [i.e., from Phe (6) formation to the monolignols 1–5 and ultimately lignin] has received recent growing interest as regards the extent of gene families/metabolic networks associated with lignification and other monolignol 1–5 related metabolism. This could thus now be fully investigated given that the *Arabidopsis* genome was sequenced in its entirety in 2000 (79). Furthermore, since several of the enzymatic steps [e.g., to cinnamic acid (8) (via PAL)] can also involve other unrelated phenylpropanoid-forming processes (such as to flavonoids, suberins, and other non-monolignol metabolites), it was therefore useful to determine to what extent there were functionally redundant metabolic networks and/or nodes within same that were monolignol/lignin-specific. This was necessary to eventually understand how these processes were (constitutively) occurring in a coordinated manner with other biochemical systems (e.g., to the cell wall biopolymers, cellulose, hemicelluloses, etc.), including identification of the various transcription factors involved.

In this regard, although possible gene families have been provisionally annotated in the TAIR database (e.g., as COMTs, CADs, etc.) (93), it became clear early on in our studies (56, 94, 122) that unambiguous biochemical/physiological functional determinations for each candidate gene (mutant) were urgently needed. Moreover, given, for example, the substrate versatility of various dehydrogenases, comparative kinetic data was also required. Thus, each of the biochemical steps discussed below had to be reconsidered in terms of the above, with particular attention being given to database (mis)annotations.

In apparent agreement with previously deduced regulatory roles for carbon allocation to the monolignol-forming pathway for C4H and *p*C3H (together with their associated reductases and HCT) described earlier, each exists in *Arabidopsis* as single gene. Moreover, while as discussed earlier, F5H does not appear to affect overall carbon allocation (77), it nevertheless serves as a “switch” between G and S lignin formation, and it is also part of a very small gene family (~2). Interestingly, until recently, enzymes such as C4H were considered to be of very limited substrate versatility. On the other hand, a recent study using recombinant C4H protein from *Arabidopsis* described a broader substrate versatility than previously recognized – albeit with unnatural substrates (146).

By contrast, the two preceding biosynthetic steps – leading to arogonate (37) and Phe (6) formation, are encoded by multi-gene families (e.g., 6 genes for the ADT and 4 for the PAL families in *Arabidopsis*). Furthermore, at least for the PAL gene family, the corresponding genes have quite complex patterns of expression [as demonstrated by GUS-reporter

promoter fusion analyses (Kim *et al.*, manuscript in preparation)]. That is, the level of co-expression in various tissues and organs is indicative of a partially overlapping, functionally redundant, metabolic network. This is anticipated to be the case for the ADT gene family as well. Both enzymatic steps also have quite strict substrate specificities, at least for the known/tested substrates employed to date.

The remaining downstream enzymatic steps (i.e., 4CLs, CCOMTs, COMTs, CCRs, and CADs) in the monolignol/lignin-forming pathway are – as originally/currently annotated in the TAIR database – members of rather large gene families, i.e., original annotations were 4CL (14 genes), CCR (12), CAD (17), CCOMT (5), and COMT (17), respectively (93). However, this is, in fact, a considerable overestimation as discussed below.

As indicated earlier, 4CL corresponds (at most) to a four-member gene family (122), and as for PAL, the GUS-promoter fusion studies for each display a complex, partially overlapping, metabolic network (Kim *et al.*, manuscript in preparation), i.e., in agreement with the different phenylpropanoid pathway products (from flavonoids to lignins) being formed. Additionally, there appear to be only 2–3 CCRs (130, 131) (Cochrane *et al.*, unpublished results) and predominantly 2 CADs (56, 57, 71) in *Arabidopsis* – at least as far as lignification is concerned. Interestingly, just one of the 17 putative COMTs has the requisite catalytic activity *in vitro* (Zhang *et al.*, manuscript in preparation) for the monolignol-forming pathway, and the CCOMTs also appear to only have at most 2 genes associated with monolignol/lignin deposition (Takahashi *et al.*, manuscript in preparation). As for PAL and other genes, the constitutive patterns of gene expression have also been established, thus providing the basis for probing/investigating the nature of the metabolic networks operative at all stages of plant growth and development.

Taken together, the entire suite of genes excluding transcription factors possibly involved in monolignol/lignin deposition – at least to the monolignols – accounts for <25 genes from prephenic acid (36) onwards, rather than the more than 10-fold higher number originally annotated.

The opportunities and challenges that now remain in further understanding monolignol/lignin pathway metabolism are to complete the full biochemical/physiological characterization of these different gene family members and to establish the “rules” for overall pathway induction/deployment – particularly since the pathway genes are “scattered” over different chromosomes. This potentially appears to be a very straightforward task, given that each of the members has now been identified – at least through their *in vitro* biochemical conversions. The data also underscore the need for circumspection when considering database annotations: while useful repositories, these cannot be considered definitive in any way until the physiological and biochemical functions are unambiguously determined.

7.5 Inherent shortcomings in lignin analyses: a critical juncture and the urgent need

Although this chapter attempts to give justice later to what we now know about native lignin macromolecular configuration and modulation thereof, it must be robustly stressed that the analytical methodologies currently available for the study of polymeric lignins (structure, content, and composition) are woefully inadequate, as well as being unsatisfactory from any

holistic quantification perspective. Indeed, such inherent shortcomings hold for all of the commonly used protocols – including the ones that we use, such as either NMR, thioacidolysis, alkaline nitrobenzene oxidation, or Klason analyses, to give just a few examples. For these reasons, it is important to appreciate and understand the very limited information that is being gained with all these approaches. As discussed below, often only a very small fraction (i.e., 20–40%) of the lignins can be accounted for in a chemically definitive and quantitative manner through degradative analyses.

Such massive deficits and/or discrepancies in the analytical procedures have, in one way or another, existed for nearly a century. For this reason, it is imperative that in the near future new approaches/technologies be urgently developed to determine, for example, the nature and frequency of *all* interunit linkages in lignin(s) – an essential requirement if we are to fully understand how lignins are, in fact, being formed *in vivo*.

While these limitations are, in large part, due to both the intractable and polymeric nature of native macromolecular lignins, there is yet another confounding feature. That is, with one exception (77), most of the research carried out on lignification (and its modulation) to date has *not* been directed toward any systematic and holistic determination of trends in the lignin-forming processes. Furthermore, there are increasing numbers of reports of lignin contents and compositions that are unreliable, and which presumably reflect a lack of scientific rigor and knowledge about lignin/cell-wall chemistries. We would plead that reliable/rigorous chemical and biochemical analyses must be an expectation for lignin investigations as they are for all other fields of (plant) chemistry/phytochemistry.

An additional complicating feature in the analysis of native lignin macromolecular configuration is that of the large variations in H, G, and S monomeric compositions and lignin contents between and within plant species. Quite remarkably, no comprehensive attempts to correlate such variability in lignin content/composition with cell wall anatomy have been reported. Yet, this would appear to be useful in terms of potentially identifying simplifying/unifying features in lignin macromolecular configuration, e.g., by correlating lignin amounts/composition with overall dimensions of the vascular apparatus and distinct cell types present in different plant lines, such as in the syringyl-rich fiber cell walls, for example.

7.5.1 Lignin isolation procedures

In spite of their intractable nature, lignins can be fully removed, but only under quite drastic chemical conditions using either strong acid [e.g., sulphite pulping processes (147–150)] or alkaline [e.g., kraft (8)] conditions at elevated temperatures and pressures. They can also be *partially* removed using milder procedures, such as by extensive ball milling for 4–5 days followed by extraction in dioxane:H₂O (9:1) (71). Indeed, the “milled wood lignin” or “Björkman” lignin (151, 152) preparations are often considered as representing the mildest forms of treatment necessary to solubilize lignin-derived components with these frequently being described as structurally closest to native lignins. The overall yields resulting from such ball-milling manipulations are though generally low ($\leq 20\%$), but can be somewhat improved by pretreatment with cellulase(s) and other hemicellulose degrading enzymes. There is also one report of attempting to solubilize entire plant tissue using a mixture of DMSO/tetrabutylammonium fluoride (TBAF) on very finely ground plant cell wall residue (CWR) (153). In our hands, this approach was unsuccessful, in large part because total

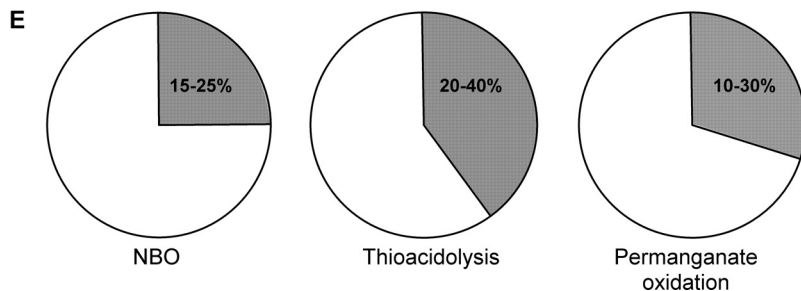
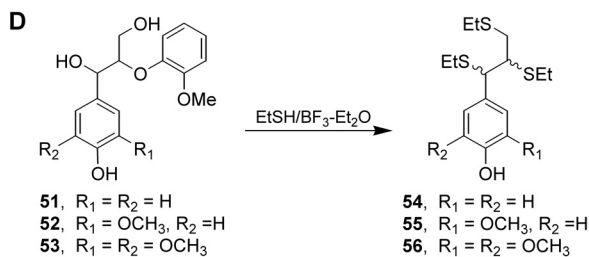
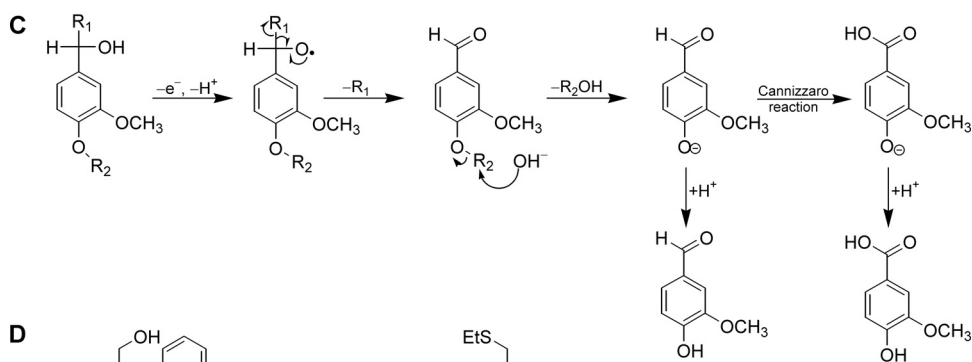
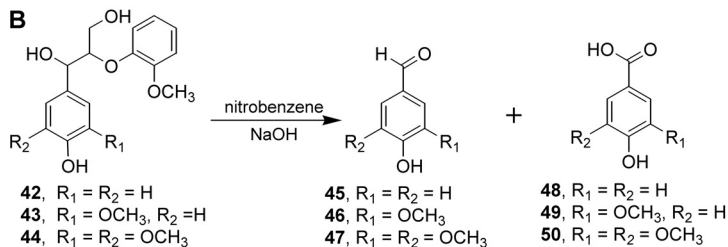
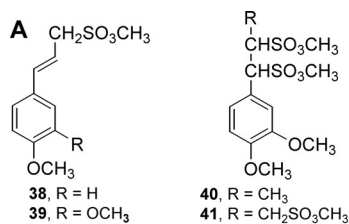
solubilization was not achieved, with the suspension also becoming very black, viscous, gelatinous, and unworkable.

In terms of the acidic sulphite treatment of woody plant tissue, the effects on delignification have been somewhat studied. For example, using a continuous flow (pulp) apparatus, it was possible to essentially delignify the tissue completely when treating black spruce (*Picea mariana*) stem wood. Characterization of the components in the resulting eluant established that initially various paucidisperse low molecular weight moieties, such as mono-, di-, and tri-sulphonated monomers, e.g., **38–41** (Figure 7.10), were released and these were readily identified (147–150). These were then followed by solubilization of higher molecular weight sulphonated polydisperse lignin chains of increasing molecular weights ranging from 3.3 to 120 kDa (147); however, these polymeric entities have not yet been fully characterized in terms of their chemical structures thus far.

Alkaline (kraft) pulping delignification has also long been investigated, together with the nature of the accompanying chemical delignification reactions. Perhaps most interesting, Sarkanen *et al.* (8) demonstrated that solubilized “kraft” lignins, assumed to contain reflections of the original lignin primary structures, displayed a capability to self-associate. This was interpreted then, and now, as due to extensive noncovalent dynamic electronic correlations between the associating (primary) chains resulting in aggregation of the polymeric lignin chains (e.g., up to several millions in molecular weight). This property is yet another complicating and potentially confounding feature in lignin analyses; such properties limit further the ability to readily study lignin from a structural perspective (e.g., by NMR spectroscopy) as discussed below.

7.5.2 Lignin subunit and lignin structural analyses by NMR spectroscopy

For a period now approximating well over 30 years, NMR spectroscopy has been applied to the study of various lignin isolates (71, 131, 132, 154–177). In this way, it has been possible, together with the study of various dimeric monolignol-derived products, to identify a number of the most abundant substructures in lignins. Many of these were previously shown in Figure 7.2D, for coupling of the H-, G-, and S-derived monolignols, respectively, with the so-called 8–O–4' interunit linkages generally being acknowledged as the most prevalent. There are, however, still major limitations in current NMR spectroscopic analyses. One is in the inability to determine the *sequences* of interunit linkages within the biopolymers, because it is not yet currently possible (using natural abundance ^{13}C) to readily go beyond the ether interunit linkages (e.g., 8–O–4', 4–O–5' linkages) to adjacent flanking substructure(s). A second major difficulty is that of the polymeric nature of the lignins: in general, the spectral band width lines for polymers can be very broad, due to molecular weight (size), polydispersity of samples, molecular aggregation and molecular heterogeneity, etc. For high molecular weight entities, such as with lignins, the molecules can thus experience slow tumbling which, in turn, results in very large transverse magnetization rates (efficient T2 relaxation) leading to broad spectral band width lines. Furthermore, given that all lignin isolation procedures result in preparations that are polydisperse, this – together with possible molecular heterogeneity (= different substructures) in the polymeric chains – can lead to



very small changes in chemical shifts at the local bonding sites and hence signal broadening. A further complication is the ability of the lignins to aggregate/self-associate which further confounds the spectroscopic analyses.

On the other hand, it is also well known that the more mobile functionalities in polymeric backbones can readily be detected [e.g., acetate groups of xanthan gum preparations (178)], whereas the polymeric backbones are more extensively line-broadened. A similar situation also presumably holds for different segments of the lignin polymer chains. Furthermore, even with observable resonances by various forms of NMR analyses, quantification of the signals relative to the entire polymeric entities can be quite problematic, again further illustrating issues as regards precise quantification. As a further caveat, chemical degradation protocols can be used to identify, detect, and quantify various lignin substructures which cannot readily be detected by NMR spectroscopic techniques. Thus, NMR spectroscopy, while a most powerful technique, currently only provides a very incomplete assessment of the nature of the lignin macromolecule(s). For these reasons, the question of lignin structural analysis has thus been limited to date in attempts to identify both interunit linkages, and less precisely to estimations of probable amounts of the distinct substructures, in the various lignin-enriched isolates – at least, for those that can be distinguished/identified/detected.

7.5.3 Quantification of lignin amounts, lignin degradation protocols, and synthetic dehydropolymerizates

7.5.3.1 Klason, acetyl bromide, and thioglycolic acid estimations

Of these three procedures, the two most commonly used for estimating gross lignin contents are the Klason and AcBr methods, respectively. The first method, albeit very routinely applied, is most uninformative, since it only measures insoluble material remaining following “digestion” with 72% H₂SO₄ (179). While generally reliable for mature woody plant stem material, it has substantial limitations when generically applied to both herbaceous and immature woody tissues [discussed in Anterola and Lewis (77) and references therein], as well as for other tissues, such as bark. Indeed, as long ago as 1986, Leary *et al.* (180) carried out Klason lignin analyses on 15 samples of hardwoods, softwoods, and grasses; it was determined from these studies that various non-lignin components, such as tannins, were present in the “lignin” isolated thereby making the analyses of solely insoluble material suspect for various sample types.

← **Figure 7.10** Sulphonation, alkaline nitrobenzene oxidation, and thioacidolysis products. (A) Sulphonated lignin-derived monomers **38–41** (147). (B) Products formed by alkaline nitrobenzene oxidation of 8–O–4′-linked model compounds **42–44**. (C) Proposed mechanism for NBO lignin cleavage. [Adapted from Schultz and co-workers (189, 190).] (D) Thioacidolysis monomeric products **54–56** from 8–O–4′ model compounds **51–53**. (E) Maximum amounts (and ranges) of monomeric/dimeric products from lignins that are typically released by alkaline nitrobenzene oxidation (NBO), (71, 132) thioacidolysis (71, 132) and permanganate oxidation; the bulk of the lignin biopolymer is largely unaccounted for.

Two recently published examples of Klason lignin contents in juvenile poplar and tobacco stems serve to further illustrate the current state of disarray in acquiring reliable basic analytical data: A report of “lignin” contents in immature (approximately 1-year old) poplar (*Populus tremula* \times *Populus alba*) stems gave values as high as $\sim 32\%$ (181), whereas others analyzing 3-month and 2-year-old poplar had indicated that levels were $\sim 20\%$ (182–184). Values of $\sim 32\%$ are well outside the ranges expected, since it is known that, for example, mature poplar wood tissues only have lignin levels ~ 18 – 21% (5). If lignin contents can be overestimated by up to nearly 60%, such approaches are unlikely to identify meaningful trends in lignin deposition/composition and assembly proper, as well as on the kinetics of lignin removal. Another quite similar example also occurred in tobacco stem analyses (69), whose sections were reported to contain ~ 40 – 50% lignin rather than the 20 – 25% expected. Both examples reflect simply a degree of unreliability in the data obtained through Klason lignin analyses, and thus a departure from the analytical rigor expected.

Similar concerns about the unreliability of thioglycolic acid lignin determinations have also been noted and critically evaluated (77). Two examples will again suffice: thioglycolic acid levels of presumed lignin contents in stem sections of 4CL downregulated *Arabidopsis* stems were considered to be $\sim 50\%$ lower than that of wild-type levels (185). On the other hand, reanalysis of this study by Anterola and Lewis (77) established that their alkaline nitrobenzene oxidation protocols had given $>115\%$ recoveries of lignin-derived fragments, rather than the $\sim 25\%$ or so expected, and suggest unreliability in one or both procedures. These results again reflect significant departure from the data expected. A similar level of unreliability was also noted for COMT downregulated alfalfa analyses: lignin contents were considered to be reduced by $\sim 50\%$ (186), although many other studies (77) have demonstrated that there were no significant levels of reduction in lignin amounts following COMT downregulation.

The AcBr method “solubilizes” lignins, as well as other non-lignin phenolics, as brominated derivatives; it has also been applied to numerous lignin determinations, using a generic extinction coefficient of $\epsilon_{280} = 20.09 \text{ L g}^{-1} \text{ cm}^{-1}$ (179, 187). Again, while very routinely used, this method does not take into account the differences in extinction coefficients due to variations in lignin monomeric H, G, and S compositions. In our more recent investigations using H-, G-, and S-enriched lignin isolates, the *best current estimates* of the actual extinction coefficients (λ , 280 nm) were established to be considerably different, i.e., (H) 15.3, (G) 18.6, and (S) $14.6 \text{ L g}^{-1} \text{ cm}^{-1}$ for *p*-coumaryl (1), coniferyl (3), and sinapyl (5) alcohol-enriched lignins (71, 188). These differences thus again underscore the need for both circumspection and scientific rigor in the study of lignins, i.e., as routinely expected for all other areas of natural product chemistry.

7.5.3.2 Alkali nitrobenzene oxidation/thioacidolysis/permanganate oxidation degradation procedures

Alkaline nitrobenzene oxidation (NBO), thioacidolysis degradation, and permanganate oxidation procedures are also currently routinely applied to the analysis of plant materials, due to their abilities to cleave various linkages in lignins, as well as that of related phenolics. For example, NBO oxidation of lignin “model” compounds **42–44** results in formation of the

various lignin-derived products (45–50) (Figure 7.10B) (189, 190), and is thus employed to probe both lignin compositions and contents. With lignins, this results in homolytic oxidative fission of their 7–8 linkages, and ultimately cleavage of the 8–O–4' bonds (Figure 7.10C) (189, 190). Moreover, if there are significant amounts of other non-lignin cell wall bound *p*-hydroxycinnamic acids/aldehydes present, this method gives overestimations of lignin contents/compositions (77). Thioacidolysis is also now perhaps even more routinely employed, again to probe both lignin contents and composition; with this method, the overall main monomeric, monolignol-derived, cleavable degradation products, obtained from lignins proper are compounds 54–56 (Figure 7.10D). These conversions have also been well studied with model compounds 51–53 that contain 8–O–4' interunit linkages. Other work is currently under way to fully identify dimeric and other oligomeric entities released by this method. Both NBO and thioacidolysis procedures are generally also employed to estimate H:G:S ratios, as well as amounts of releasable products relative to lignin contents and/or cell wall residues (CWR). Overall, the recoveries of monomeric and dimeric components, relative to original lignin contents, are very low. For example, alkaline NBO and thioacidolysis generally only account for *circa* 15–25% and 20–40% by weight, respectively, of the lignin presumed present (71, 132) (Figure 7.10E). Thus the bulk of the lignin polymer(s) remains unaccounted for using either procedure. Lastly, permanganate oxidation has also been quite extensively employed with various carboxylic acid methyl esters ultimately being formed (191). The yields of these products are also low (192) (~10–30% or so, Figure 7.10E), with the bulk of the lignin polymer again being unaccounted for.

In spite of these serious shortcomings, as discussed below, little has been done until now to attempt to establish overall trends in interunit linkage placement (as a function of lignin deposition); in accurately determining response factors, etc., for calibration and quantification purposes; in developing new/improved methodologies to study both lignins and their degradation products. For these reasons, and before discussing the results so obtained, it was thus first necessary to consider the advantages and limitations of the procedures currently employed.

7.5.3.3 Synthetic dehydropolymerizates

Synthetic dehydropolymerizates (DHPs, so called “lignin-like”) can be obtained by H_2O_2 /peroxidase catalyzed one-electron oxidation of monolignols. While generating some of the substructures found in native lignins, there is, of course, little opportunity for control over primary chain interunit linkage (sequence) designation with *in vitro* preparations. Thus, numerous important differences between synthetic DHPs [which, for example, release dimeric pinosresinol-derived substructures (see Figure 7.2D, substructure **IIIb**) upon cleavage] and native lignin polymers [which do not (193)] have long been known, and which diminish greatly their utility (29). While it is possible to synthesize 8–O–4' linked oligomers (194), and also to modulate the amounts of 8–O–4' interunit linkages through varying peroxidase activity *in vitro* (195), such preparations are not lignins proper. Indeed, in many respects, the comparison of synthetic DHPs to native lignins might be likened to that of randomly linking amino acids through amide linkages in the same relative proportions as found in insulin, and describing the resulting products as an insulin mimic.

7.6 Modulation of monolignol pathway and peroxidase enzymatic steps: predictable effects on the vascular apparatus and on limited substrate degeneracy during proposed lignin template polymerization

Increasing demands for renewable resources, combined with modern biotechnology capabilities, have driven recent efforts to “improve” commercially important plant-derived products. In this regard, reduction of total lignin content and alteration of lignin composition are currently major goals of research in phenylpropanoid metabolism, particularly to now help overcome the challenge of recalcitrant lignocellulose. While considerable effort continues to be directed toward such goals, little consideration has been given to side effects to either plant (anatomical) structure and/or plant defense systems. One exception to this is the article by Anterola and Lewis (77), which began the first thorough review of the effects on plant structure, and this section builds on the knowledge and insights gained. In this regard, the effects of lignin and its downregulation on plant structure are still not well understood for a variety of reasons including: (a) little to no detailed observation on plant structure; (b) variable and non-comparative results reported because of unreliable methods of mutational/transgene generation and/or lignin analyses employed; (c) lack of knowledge and often unfounded assumptions regarding the physiological roles of individual isoforms of (presumed) phenylpropanoid enzymes; (d) assumptions that abnormal monomers can compensate for monolignol downregulation or ratio alteration; and (e) a very limited understanding of how plant physiological processes/architecture are modulated and the factors associated with same.

In this regard, the metabolic flux studies described earlier (34, 35) had provided very useful insight into rate-limiting processes in monolignol/lignin-forming pathways and which ultimately control carbon allocation to the H and G(S) segments, i.e., through regulation of Phe (6) formation and by differential control of C4H and *p*C3H enzymatic activities. These insights thus provided a beginning for formulating models for predicting the consequences/effects of altering lignin biosynthesis, whether with various transgenic lines or mutants of the monolignol/lignin pathway. Indeed, they now provide a scientific basis to explain results previously obtained with various lignin pathway mutants – some going back as far as to the early 1930s. Additionally, the approach here has facilitated examination of the question of non-random lignin assembly, lignin macromolecular configuration, and effects of modulating same via *limited* substrate degeneracy during proposed template polymerization.

The need for this systematic reasoning and the corresponding analyses in the above areas contrasts markedly with that of some other research endeavors. A very different paradigm, largely attempting to explain the effects of transgenic/mutant organisms obtained, has instead proposed a “random coupling/combinatorial chemistry” model to account for presumed findings and/or interpretations made (175). Beginning with the notion that native, monolignol-derived, plant lignins [say of a 100-dp (mer) chain length] can have an astronomical number of different isomers (i.e., 10^{66}) (196), this was later extended in a section entitled “Personal Ranting” (197) to espouse the view that “the actual composition of lignin is not particularly important.” Relevant sections are provided below for needed context and discussion, since the two views [our own and those of Ralph *et al.* (175, 197)] are

widely divergent in terms of considerations of both lignin assembly system(s) and lignin macromolecular configuration.

Warning. Personal Ranting Section. I suggest the blasphemous idea that the plant simply needs a polymer with required properties and that lignin's composition is not particularly significant!! ... Perfectly viable plants are obviously produced when blocking their enzymes and inhibiting the production of crucial monomers forces them to make lignins from abnormal precursors ... That plants can respond instantly to severe enzymatic shifts, inflicted on them in a single generation, without the benefit of evolution, suggests that lignin's structure is not that important. (197)

By 2004, based on the above interpretations and other research findings, it was further proposed that:

The striking realization that hits one contemplating these notions is that it becomes unlikely that there are two lignin molecules over a given degree of polymerization (DP) that are structurally identical in all the plants on this entire planet. (175)

Such reasoning could clearly obviate the need to examine lignification for trends in deposition, orderedness in structure, and assembly, etc. Indeed, as indicated earlier, the latter is urgently needed given the deficits in the analytical procedures currently available.

The scientific challenge posed by such assertions is, therefore, in how to reconcile such widely divergent and apparently incompatible views about lignin macromolecular configuration and assembly. In this part of the chapter, we thus consider it useful to briefly and systematically describe findings obtained thus far for each mutant and/or transgenic line, and to question the validity of each hypothesis in doing so. The enzymatic steps that have been modulated thus far (i.e., with transgenic and/or mutated plant lines) and which help address such questions are PAL, C4H, *p*C3H, HCT, 4CL, CCR, CAD, COMT (CCOMT), F5H, peroxidase, and laccase. In terms of the two oxidases (peroxidases and laccases), attention is also given later to dirigent proteins and proteins *proposed* to harbor dirigent (monolignol radical) sites in Section 7.7. Additionally, a brief discussion of transcription factors controlling secondary wall formation in fibers is given (198, 199). It must be emphasized at the beginning, though that all these investigations, while of scientific interest, represent very preliminary studies. More importantly, they underscore our current paucity of understanding of plant metabolism, growth, and development, and of plant cell wall formation and lignin primary structure in general.

7.6.1 PAL, C4H, *p*C3H, HCT, and 4CL downregulation/mutation

7.6.1.1 PAL

This step catalyzes the entry point into phenylpropanoid metabolism, and thus can lead to a diverse range of phenylpropanoid-derived products (i.e., lignins, lignans, hydroxycinnamic acids, suberins, flavonoids, proanthocyanidins, etc.), depending upon the species, tissue, and/or cell type in question. Reduction of *overall* PAL enzymatic activity (through inhibition, mutation, downregulation, etc.) would thus be expected to have negative consequences on either all aspects of phenylpropanoid metabolism or on specific elements, depending upon

which *PAL* gene(s) is (are) affected. This was facilely demonstrated first using the *PAL* inhibitor, L-AOPP (57, Figure 7.11), as early as in 1977 and in subsequent studies extending through 1985 (200–205). This resulted, depending upon the plant species investigated, in reductions in formation/accumulation of anthocyanins, isoflavones, hydroxycinnamic acids, and lignins, respectively. As an example, Figure 7.12A depicts the effects of AOPP (57) inhibition on lignification in mung beans (2004); the cross-section of the AOPP (57)-treated plant line clearly has collapsed xylem due to presumed reductions in lignin contents and thus a weakened vasculature – as would be anticipated from first principles.

Furthermore, although the reliability of many of the analytical techniques used in later studies by other researchers was clearly questionable, the subsequent standard biotechnological manipulation of *PAL* activity in tobacco (through antisense and sense co-suppression, etc.) apparently resulted in poorly formed (weakened) xylem tissue (206, 207); additionally, the plant lines were also severely stunted, had curled leaves, localized lesions, less pollen, reduced viability and deformed flowers. These are presumably *not* desirable traits, and indeed appear to fall under the rubric of pleiotropic “unintended” effects: these, in turn, reflect our severely limited current understanding of overall plant metabolism, growth, and development. Moreover, while the lignin levels were reportedly lower in these studies as indicated above, the analytical procedures used were very questionable (i.e., thioglycolic acid lignin determination and Klason lignin analyses of “neutral detergent fibers”) as discussed in Anterola and Lewis (77). Another very preliminary study also reported that generation of double *PAL* mutants (e.g., *pal1 pal2*) in *Arabidopsis* resulted in lignin levels reduced to *circa* 30–35% of wild-type levels (208). This study, however, lacked any systematic analysis of the effects on lignification over different phases of growth/development until maturation. It did though provisionally indicate that lignin levels were reduced.

Perhaps, most importantly, none of the above studies explored any (quantitative) measurement of effects on structural integrity of the vasculature; nor was there either any new or additional insight gained on lignin macromolecular configuration/assembly, and/or effects of modulation of same. As predicted from first principles by ourselves there was clearly no “instant response,” as proposed by Ralph (197) and Ralph *et al.* (175), to severe enzymatic shifts that would suggest that lignin’s structure was not important; nor were “perfectly viable” plants produced, based on the defects that had been noted as early as 1983. That is, the genetically modified tobacco *PAL* plant lines produced were only able to survive with a number of serious defects; additionally, these transformants were not stable, and reverted to wild type in subsequent generations (209). Accordingly, such preliminary (phenomenological) studies on *PAL* modulation now need to be expanded in depth, in order to gain a more definitive understanding of the actual effects on, for example, lignin macromolecular configuration/assembly, vascular integrity and overall physiology.

7.6.1.2 C4H

A similar situation to that for *PAL* downregulation/mutation holds for *C4H*. Although this enzyme is considered to be one of the important rate-limiting processes in phenylpropanoid metabolism (such as for differential carbon allocation) (34, 35), it is also a common step leading to various metabolites, including monolignols, lignins, flavonoids, and so forth. Downregulation of *C4H* has though only been studied in a very preliminary way thus far

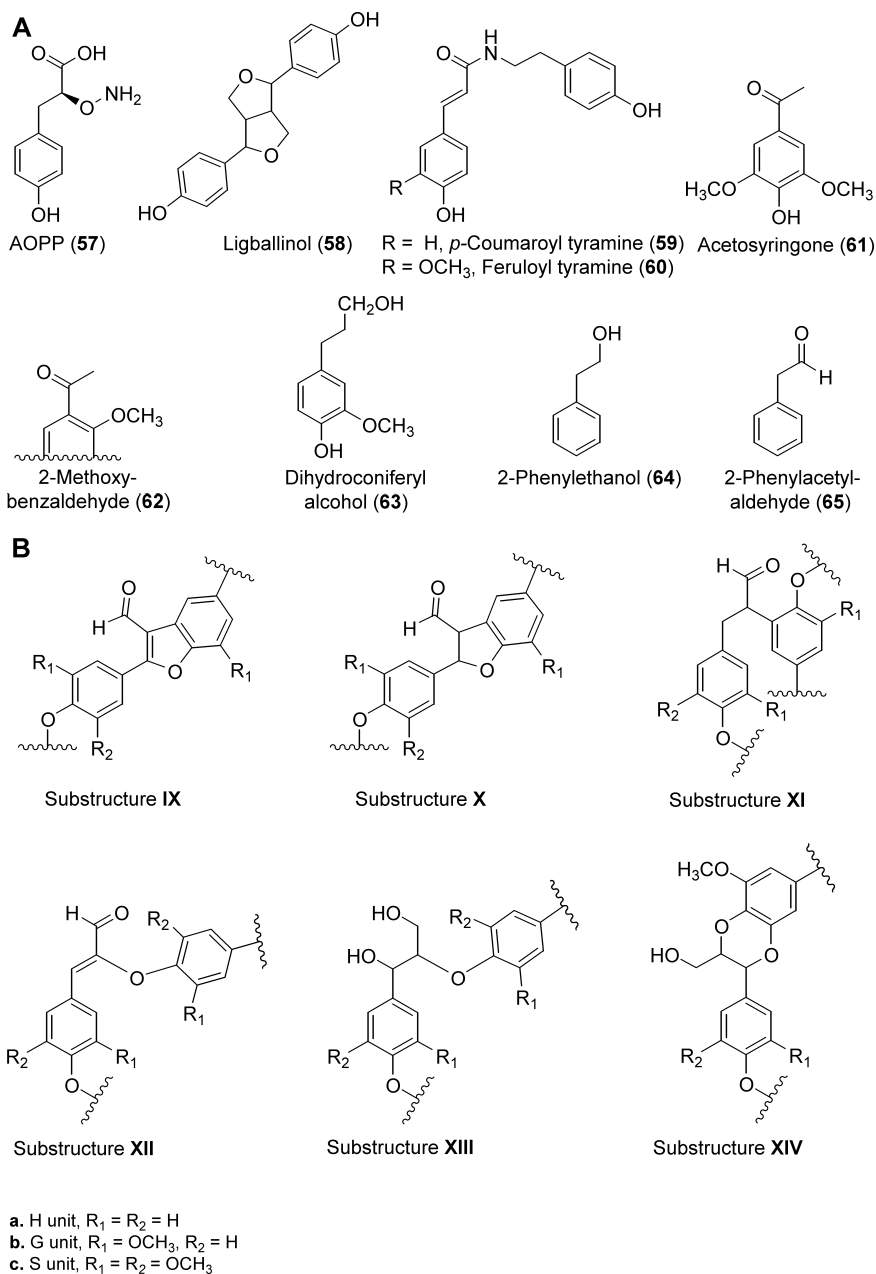
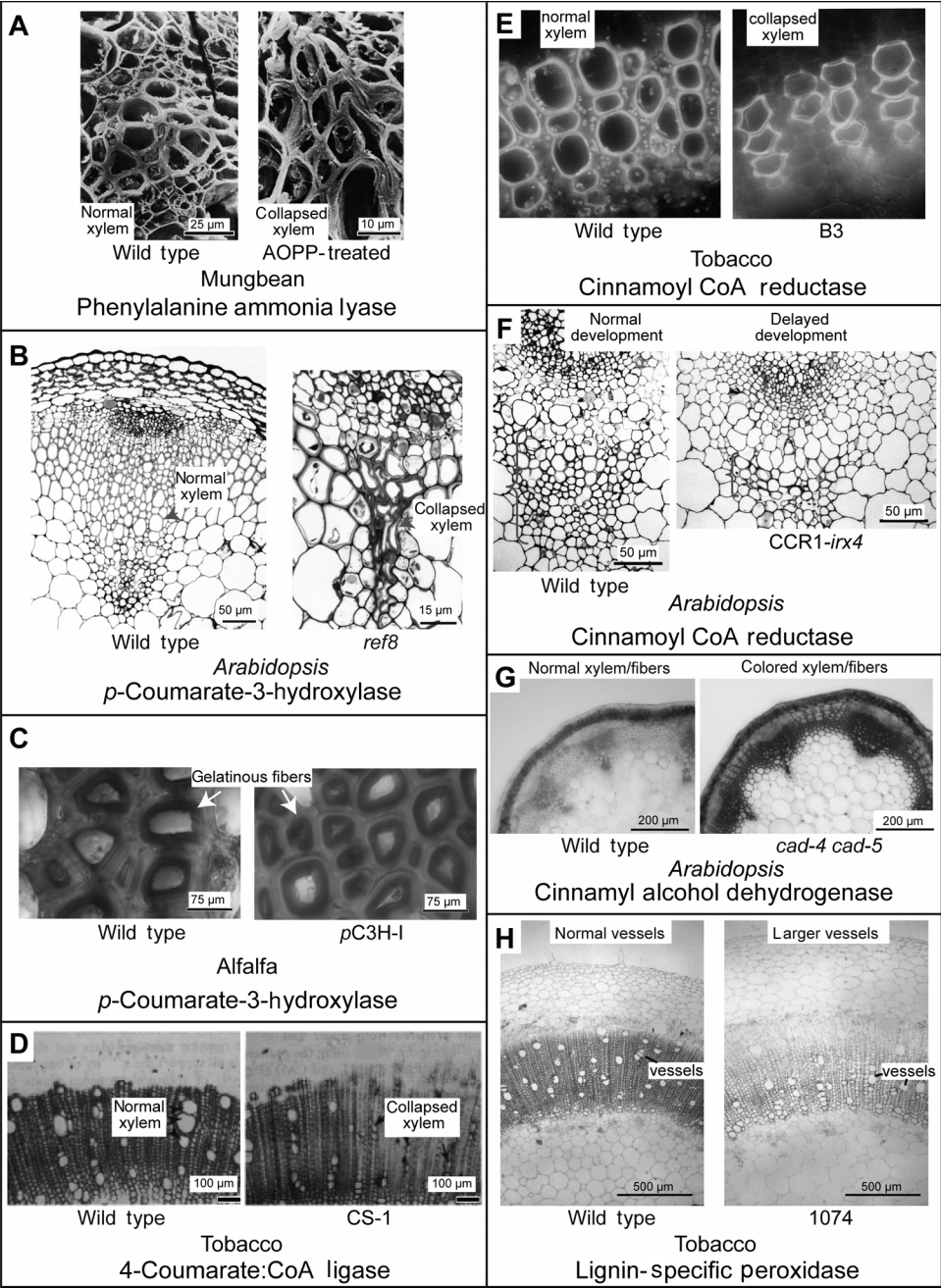


Figure 7.11 PAL inhibitor, AOPP (57), various monomeric/dimeric phenols and aromatics 58–65 and (B) various phenolic-derived substructures.



(77, 207, 210–212) as regards overall effects on lignification and the vasculature. As noted above for PAL, the methodologies employed in two of these studies (207, 212) were again very questionable (i.e., estimations of Klason lignin of neutral detergent fiber), and ultimately found to be unreliable as analytical protocols. Nevertheless, in all three studies the lignin levels provisionally appeared to be reduced – at least by >25% (207, 210, 211). Various alterations were also noted in S/G ratios of the lignified tissues, the reasons for which are yet unknown. As for PAL modulation, no evidence was obtained indicating deployment of non-monolignol moieties to “compensate” for reductions in lignin levels. There were also no experiments conducted to quantitatively ascertain the effects on the vascular integrity through biomechanical analyses. In hindsight, such studies need to be expanded significantly in order to ascertain, at the very least, the true effects of reductions in lignin levels on plant growth/development, on overall vascular structure/anatomy, and on lignin macromolecular configuration, etc.

Figure 7.12 Examples of effects on vascular anatomy of inhibiting/mutating/downregulating various enzymes and/or genes in the phenylpropanoid pathway. (A) AOPP-treatment of mungbean (204), results in collapsed xylem. (B) Similar effects occurred in a *pC3H* mutated *Arabidopsis* line (*ref8*). (C) On the other hand, formation of gelatinous fibers such as in wild-type alfalfa (*Medicago sativa*) apparently help mechanically offset deleterious effects of *pC3H* downregulation without forming “abnormal” lignins (72). (D) Collapsed xylem observed in tobacco (*Nicotiana tabacum*) lines 4CL downregulated (223), as was (E) xylem in a tobacco line CCR downregulated (232). (F) Detailed analysis of the *Arabidopsis irregular xylem4* mutant (CCR1-*irx4*) identified pleiotropic effects, including delayed xylem formation (131). (G) An *Arabidopsis* CAD double mutant, AtCAD4/5, (*cad-4*, *cad-5*) with unusual red coloration in the xylem attributed to presence of *p*-hydroxycinnamyl aldehydes (71). (H) Qualitatively larger vessels observed in a tobacco line downregulated for a lignin-specific peroxidase (257). All images are of transverse sections and taken using brightfield microscopy, unless otherwise specified. Images shown in (A) are SEM analyses, while images (B and F) are images of semi-thin resin embedded sections stained with either toluidine blue O or Stevenel’s Blue. Image (C) was taken of cryosections stained with zinc chloro-iodide. Image (D) was taken of hand-cut sections stained with phloroglucinol-HCl. Image (E) was taken using UV light microscopy. Image (G) was recorded using unstained hand-cut sections, while image (H) was taken of phloroglucinol-HCl stained sections. [Reprinted from (A) The European Journal of Cell Biology, vol. 29, Amrhein, N., Frank, G., Lemm, G. & Luhmann, H.-B., Inhibition of lignin formation by L- α -aminooxy- β -phenylpropionic acid, an inhibitor of phenylalanine ammonia-lyase, pp. 139–144, Copyright 1983, with permission from Elsevier. (D) Plant Science, vol. 128, Kajita, S., Mashino, Y., Nishikubo, N., Katayama, Y. & Omori, S., Immunological characterization of transgenic tobacco plants with a chimeric gene for 4-coumarate:CoA ligase that have altered lignin in their xylem tissue, pp. 109–118, Copyright 1997, with permission from Elsevier. (E) Plant Journal, vol. 13, Piquemal, J., Lapiere, C., Myton, K., O’Connell, A., Schuch, W., Grima-Pettenati, J. & Boudet, A.-M., Downregulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants, pp. 71–83, Copyright 1998, with permission from Blackwell. (F) Phytochemistry, vol. 66, Patten, A.M., Cardenas, C.L., Cochrane, F.C., Laskar, D.D., Bedgar, D.L., Davin, L.B. & Lewis, N.G., Reassessment of effects on lignification and vascular development in the *irx4* *Arabidopsis* mutant, pp. 2092–2107, Copyright 2005, with permission from Elsevier. (G) Phytochemistry, vol. 68, Jourdes, M., Cardenas, C.L., Laskar, D.D., Moinuddin, S.G.A., Davin, L.B. & Lewis, N.G., Plant cell walls are enfeebled when attempting to preserve native lignin configuration with poly-*p*-hydroxycinnamaldehydes: Evolutionary implications, pp. 1932–1956, Copyright 2007, with permission from Elsevier; and (H) Phytochemistry, vol. 64, Blee, K.A., Choi, J.W., O’Connell, A.P., Schuch, W., Lewis, N.G. & Bolwell, G.P., A lignin-specific peroxidase in tobacco whose antisense suppression leads to vascular tissue modification, pp. 163–176, Copyright 2003, with permission from Elsevier.] (Reproduced in color as Plate 21.)

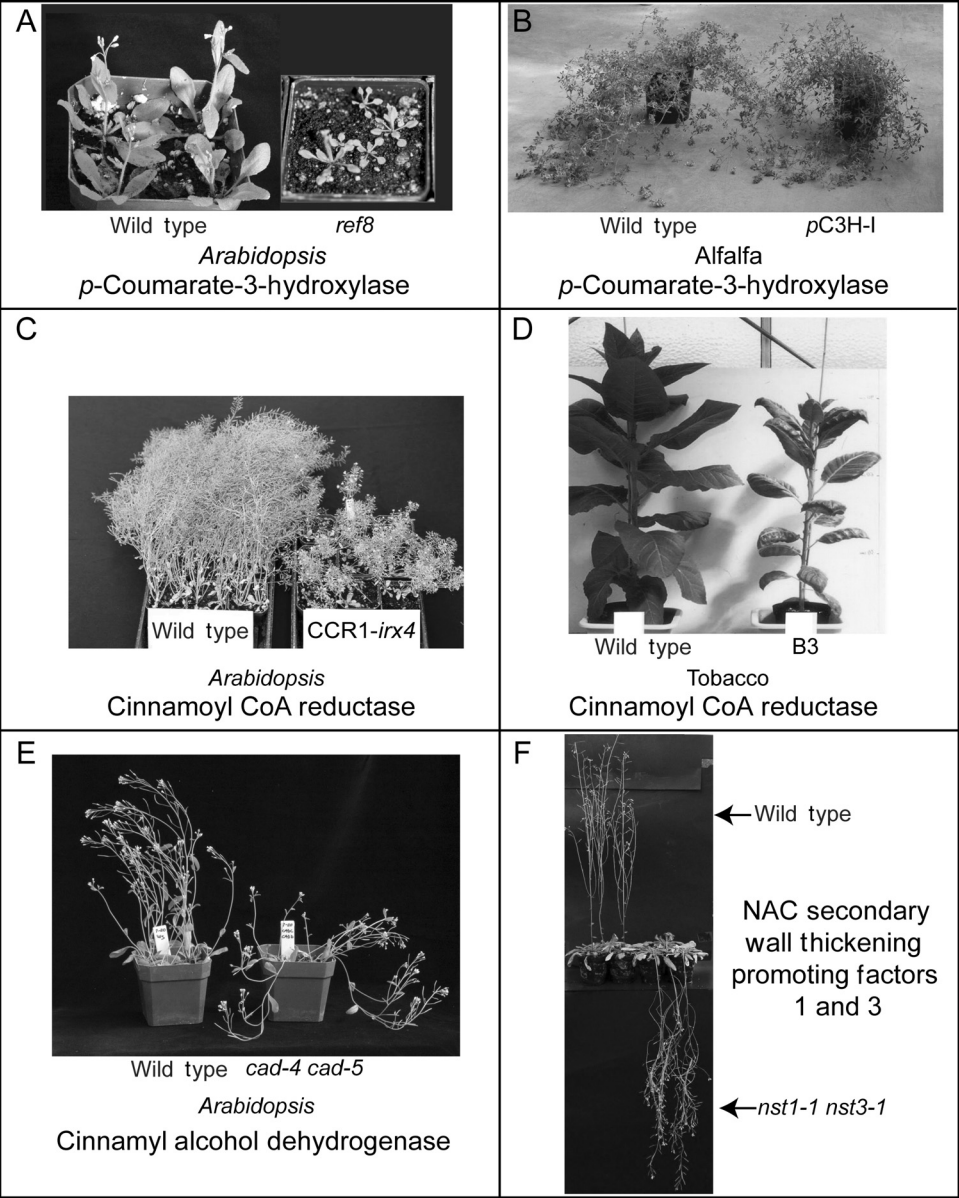


Figure 7.13 Gross phenotypical changes/effects of either mutating or downregulating various phenylpropanoid genes, as well as vascular related transcription factors. **(A)** *pC3H* mutant with various pleiotropic effects, resulted in a severely dwarfed line (*ref8*) (213, 215) (shown 23 days post seed-sown). **(B)** *pC3H* downregulation in alfalfa (*Medicago sativa*) resulted in a phenotype (*pC3H-I*) very similar to wild type without visible pleiotropic effects (shown 4 weeks post-cut-back). **(C)** Dwarfed *Arabidopsis* *CCR1-irx4* mutant with pleiotropic effects (shown 6 weeks post seed-sown) (131). **(D)** *CCR* downregulation in tobacco, with stunted morphology relative to wild type (232). **(E)** *CAD* double mutation in *Arabidopsis* (*AtCAD4/5*, *cad-4*, *cad-5*) resulted in a limp to prostrate stem phenotype (shown after 4 weeks growth) (71). **(F)** Double T-DNA tagged *Arabidopsis* *nst1 nst3-1* line (199). [Reprinted from **(B)** The American Journal

7.6.1.3 *pC3H* and *HCT*

pC3H is also considered to have an important regulatory role in carbon allocation (77) to the monolignol/lignin-forming pathway, at least as far as the G and S segments are involved. Two different mutant/transgenic lines (*Arabidopsis* and alfalfa) altered in *pC3H* activity have thus far been obtained using standard biotechnological manipulations (72, 213, 214). These manipulations resulted in one of the most dramatic examples of disparity seen in morphology between two different species, i.e., as a result of the same genetic target (*pC3H*) (72, 215).

In this regard, the chemically generated (ethane methyl sulfonate, EMS) *Arabidopsis ref8* mutant, isolated following screening of ~100 000 plants, contains a random mutation in the *pC3H* gene (213). The line was severely dwarfed (Figure 7.13A), reaching only ~3% of the height of wild type at maturity (215). It was also sterile and its proto-/metaxylem and xylary fiber elements had collapsed in the stem cross-sections examined (Figure 7.12B). This latter observation presumably reflects the very low levels of lignin, reduced by ~64% relative to wild type, and which is primarily of H character (H:G:S ratios of ~85:8:7) (215). This observation contrasts with previous assertions that the *ref8* lignin was exclusively H-derived (175, 213).

Although *pC3H* has a rate-limiting role in lignin biosynthesis and can thus be expected to produce vascular defects, the multiple and extreme morphological characteristics of *ref8* suggested that additional pleiotropic effects were manifested in this line (215). Because of tissue availability limitations, the *ref8* line was not readily amenable to more detailed analyses and thus these questions were not explored further. On the other hand, the alfalfa *pC3H-I* line was generated using a standard antisense transgenic method (214) with presumably less “unintended” effects resulting than that of chemical mutagenesis. In stark contrast to the *Arabidopsis ref8*, the alfalfa *pC3H-I* has a phenotype nearly like that of wild type (Figure 7.13B), with vessel anatomy also being very similar to wild type (data not shown) (72).

The remarkably different outcomes of modulation of *pC3H* between the *Arabidopsis ref8* mutant and alfalfa *pC3H-I* lines (215) reflect further the weaknesses of current methods in the genetic manipulation of plants, and of our very limited understanding of said manipulations. Indeed, it is for reasons such as the unknown (pleiotropic) effects observed herein that some segments of humanity currently distrust the use of genetically modified organisms. From a more scientific perspective, these limitations also highlight the need for comprehensive, careful, systematic data collection in order to more accurately assess,

←
Figure 7.13 (continued) of Botany, vol. 94, Patten, A.M., Jourdes, M., Brown, E.E., Laborie, M.-P., Davin, L.B. & Lewis, N.G., Reaction tissue formation and stem tensile modulus properties in wild type and *p*-coumarate-3-hydroxylase downregulated lines of alfalfa, *Medicago sativa* (Fabaceae), pp. 912–925, Copyright 2007, with permission from the Botanical Society of America. (D) Plant Journal, vol. 13, Piquemal, J., Lapierre, C., Myton, K., O’Connell, A., Schuch, W., Grima-Pettenati, J. & Boudet, A.-M., Downregulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants, pp. 71–83, Copyright 1998, with permission from Blackwell; and (F) The Plant Cell, vol. 19, Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K. & Ohme-Takagi, M., NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*, pp. 270–280, Copyright 2007, with permission from the American Society of Plant Biologists.] (Reproduced in color as Plate 22.)

analyze, and understand the effects of such manipulations. Clearly, we need to get such manipulations “right,” for example, before contemplating dedicating large swaths of lands to lignocellulosic-modified organisms.

The alfalfa *pC3H* line also had lignin contents significantly reduced ($\geq 68\%$ of wild type) (72, 215), which were again mainly of H-character (H:G:S ratios of $\sim 77:12:11$) as to be predicted (31, 34, 35, 77). Significantly, biomechanical testing of stem vascular tissue integrity showed essentially no differences in tensile dynamic modulus properties, at least under the test conditions employed (72), i.e., where both the *pC3H-I* and wild-type lines apparently displayed more or less equivalent structural integrity of the vascular apparatus. More detailed anatomical analyses of the stem tissue in both the *pC3H-I* and wild-type lines also established that the vasculature contained lignin-deficient, cellulose-enriched, gelatinous layers that are characteristic of reaction (tension) wood (see Figure 7.12C). Moreover, this occurred both earlier in growth/development and to a greater extent in the *pC3H-I* line (72), suggesting that vascular/structural properties are (at least in part) maintained by partially “switching” metabolic outcome to that of reaction tissue formation. Although the latter tissue is (both constitutively and inducibly) employed to control branching and realignment of leaning stem(s) back to the vertical, there are many applications, such as in forestry and in pulp/paper manufacture, where its formation is not desirable; gelatinous fibers, for example, result in poor paper strength quality due to decreased bonding properties (216).

Analyses of the lignins in both *pC3H-I* and wild-type lines were most instructive: the ^{13}C NMR spectroscopic analysis of solubilized lignin derivatives established the presence of a H-enriched lignin (215) in *pC3H-I* as predicted (31, 34, 35, 77), with substructures **I–VIII** (Figure 7.2D) being in evidence. More importantly, the thioacidolysis degradation products (monomers and dimers) released, which corresponded to cleavage products from subsets of 8–O–4', 8–5', 5–O–4', 3–3' (5–5'), and 8–1' (substructures **I/II**, **IV**, **VI**, **VII**, and **VIII** in Figure 7.2D) lignin interunit linkages, were determined quantitatively at various stages of stem growth/development/lignification until maturation (215) (Figures 7.14A, F–I). The frequency of the interunit linkages indicated that they were directly proportional to increasing lignin content, at least for the (monomeric/dimeric) cleavable 8–O–4', 8–1', 8–5', 5–O–4', and 3–3' (5–5') moieties being released. Significantly, the 8–8'-linked dimer (ligballinol **58**) was not released as such (Figure 7.14J), even though its “substructure **IIIa**” was readily detectable by NMR spectroscopic analyses. This observation is also consistent with the previous “conundrum” of pinosresinol-like substructures **IIIb** not being released either from native lignins, whether by acidolysis (193) or thioacidolysis (217). These substructure-like moieties are, nevertheless, also readily detectable by NMR spectroscopic analyses. In direct contrast, synthetic lignin DHPs readily release such entities (218) upon cleavage. This, in turn, is presumably indicative of a very different mode of lignin assembly/organization *in vivo* than that occurring randomly *in vitro*.

Moreover, when compared to the lignin levels and interunit linkage frequencies in the wild-type line, the same correlations and trends exist for all levels of lignin deposition, with the amounts of individual products released and quantified apparently being *invariant of methoxyl group composition*. Significantly, the H-monolignol (**1**) did not seamlessly replace the G/S components (**3/5**), and thus presumably was unable to extend deposition into the domains normally designated for G/S monolignols in the lignifying cell wall(s). This, in turn, indicated that the H-monolignols (**1**) presumably could not substitute fully for either coniferyl (**3**) or sinapyl (**5**) alcohols, this representing yet another constraint on

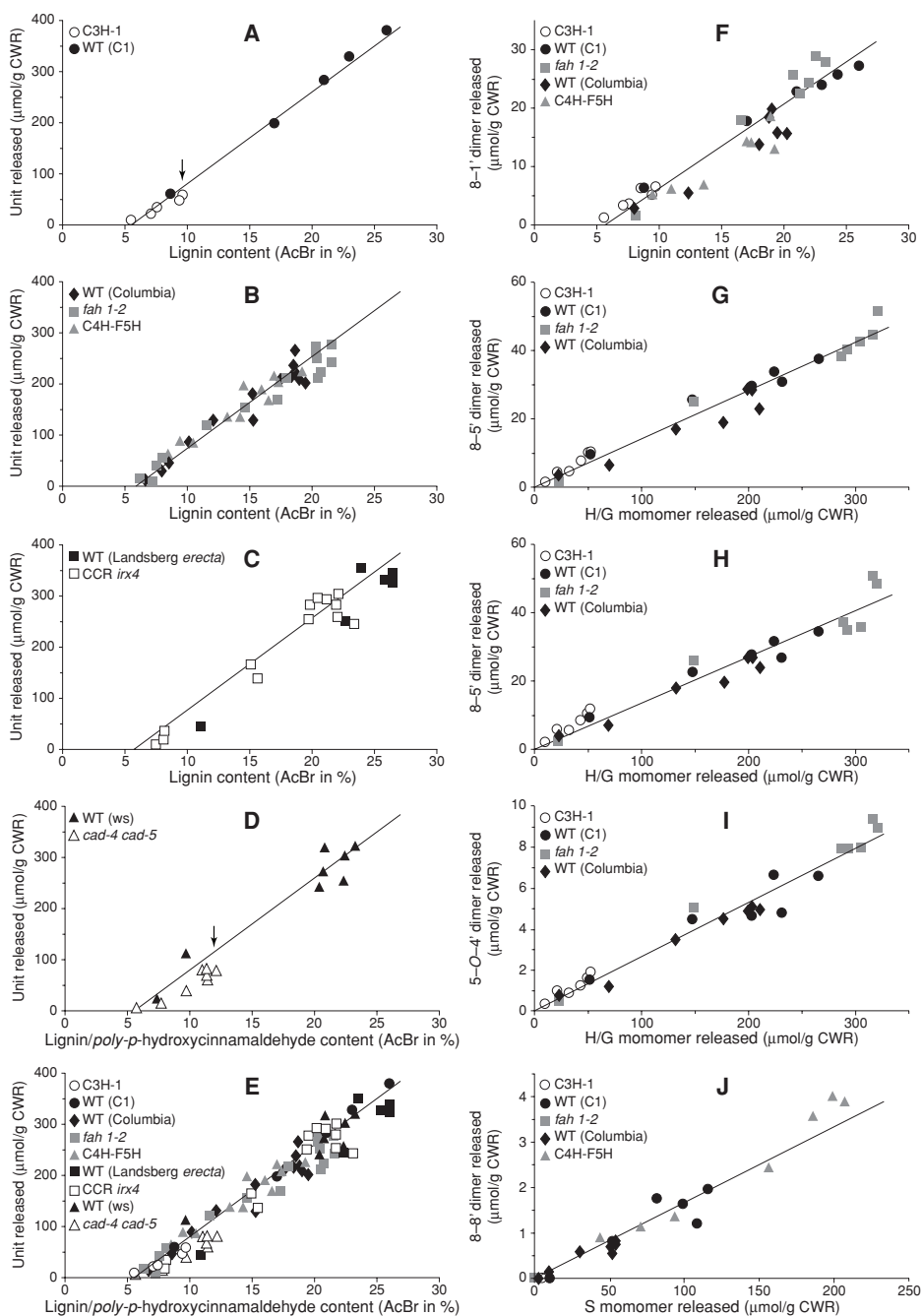


Figure 7.14 Correlations of current best estimates of amounts of (A–E) total monomeric derivatives releasable by thioacidolysis as a function of estimated AcBr lignin contents; (F) 8–1' dimeric derivatives after thioacidolysis followed by desulphurization as a function of estimated AcBr lignin contents; 8–5' (G), 5–5' (H) and 5–O–4' (I) dimeric derivatives after thioacidolysis followed by desulphurization as a function of total releasable monomeric derivatives by thioacidolysis; (J) 8–8' dimeric derivatives after thioacidolysis followed by desulphurization as a function of total S unit releasable monomeric derivatives by thioacidolysis.

the notion of a seamless random coupling/combinatorial chemistry being in effect (175). On the other hand, the overall amounts of H-moieties appeared to increase slightly over that of wild-type levels, but where designated patterns of a subset of interunit linkage frequencies (for the products released) were being maintained (relative to wild-type lignin) until the overall deposition process was prematurely terminated – for reasons still to be determined. Such reasons for terminating this process could include the plant perceiving that the biophysical properties of the H-lignin being deposited in the cell wall were, for example, limiting and/or defective in some way, e.g., through feedback inhibition. Nevertheless, the *partial* replacement of G/S moieties by H-monomers can presumably be explained via limited substrate degeneracy during lignin template polymerization.

Reduction in lignin contents have also been reported with HCT silencing in *Nicotiana benthamiana* and *Arabidopsis* (219, 220). For example, RNA silencing was achieved in *Arabidopsis* (ecotype Columbia) by transformation with a hairpin repeat of a portion of the HCT gene whereas in *N. benthamiana*, a tobacco rattle virus (TRV)-based virus-induced gene silencing (VIGS) system was used (219). In *Arabidopsis*, this resulted in stunted plants with apparently reduced lignin levels based on collapsed xylem cells (220). The yields of thioacidolysis monomers **54–56** of stems from 2-month-old plants were, however, very low, i.e., $\sim 75.7 \mu\text{mol g}^{-1}$ dry sample for the wild type and only $\sim 1.0 \mu\text{mol g}^{-1}$ dry sample for the silenced line (HCT⁻). Such values for the wild type (control) are uncharacteristically low and presumably indicate that the lines were of an unknown maturation state, and/or were defective in some manner, and/or the analyses had yielded unreliable data; in our studies and others, the thioacidolysis yields are reproducibly $\sim 350 \mu\text{mol g}^{-1}$ dry sample (cell wall residue, CWR) for wild-type lines at plant maturity (see Figures 7.14A–E). Thus, these data again underscore the need to determine at the minimum lignin contents and compositions at several stages until maturation has definitively been reached, rather than the single point analyses that have become typical of most studies in this field. On the other hand, the H:G:S ratios of 1:83:16 in wild type were now 83:12:2 in the HCT⁻ line in agreement with HCT involvement in G/S lignin biosynthesis. In *N. benthamiana*, plants ranging from no visual growth phenotype to severely stunted phenotypes were also obtained, yet with estimated lignin levels decreased only by $\sim 15\%$ of wild type in the HCT-silenced plants as measured by the Klason method, i.e., the results for both species were apparently quite different in terms of effects on lignification. Although the reasons are not yet known for these differences, *Arabidopsis* and tobacco differ in having only one HCT homologue in the former, whereas in the latter both HCT/HQT enzymes are present. Yet, HQT silencing was also studied in tomato (*Lycopersicon esculentum*) where it was shown to only affect chlorogenic acid (**26**) levels but not lignin content/composition (111).

7.6.1.4 4CL

Modulation of 4CL activity can also, depending upon the gene(s), be targeted to affect monolignol, flavonoid, suberin, and related metabolism. Standard biotechnological manipulations of overall 4CL activity have been described thus far using tobacco (221–223), *Arabidopsis* (185), and aspen (*Populus tremuloides*) (224). Very preliminary descriptions of each of the corresponding phenotypes so obtained were also reported to be dramatically different, even though lignin levels were apparently diminished to a similar extent by *circa* 50% or so, but only when 4CL activity levels were greatly reduced (down to $\sim 7\%$ of wild-type

levels) (77). The reports on the transformants ranged from having no visible phenotype in *Arabidopsis*, to dwarfing in tobacco, and to a claim of spectacularly enhanced growth in aspen.

In the case of 4CL downregulation in tobacco, the data obtained were the most comprehensive and generally appeared the most reliable (77, 221–223), albeit where several of the transformants were dwarfed at the flowering stage. In terms of the ^{13}C NMR spectra of the lignin-enriched preparations isolated from the wild type and 4CL downregulated lines, both were very similar except for a few slightly enhanced resonances which were subsequently interpreted by Anterola and Lewis (77) as being mis-assigned by Kajita *et al.* (222). We considered these to correspond instead to *p*-coumaroyl (59)/feruloyl (60) tyramine moieties, which if correct, were then in accordance with our previous work on this molecular species in tobacco (225). There was no evidence that the overall amounts of these well-known tyramine-derived moieties in the *Solanaceae* had either increased, or were covalently linked to the lignin in the lignin-enriched isolates. Interestingly, and as expected, the 4CL downregulated line displayed a significant reduction in vascular integrity as evidenced by collapsed xylem in the tobacco stem cross-sections (223) (see Figure 7.12D). Again, though, there were no comprehensive studies on tobacco growth/development and structural integrity of the vasculature of these transformants carried out to gain a more full understanding of the overall effects engendered by this manipulation.

For 4CL downregulation of *Arabidopsis* (185), the phenotype was apparently similar to wild type, although no data were provided describing either growth and/or developmental processes. Additionally, the results obtained were not explicable based on known lignin chemistry, even though ostensibly there were reductions in lignin contents. For example, the approach taken by these researchers to estimate lignin amounts used the unreliable thioglycolic acid method. Nitrobenzene oxidation (discussed earlier) was also carried out, this being a degradative method which generally accounts at best for *circa* 15–25% of the lignin present in plant tissue (Figure 7.10E). Recalculation and comparison of the thioglycolic acid and nitrobenzene oxidation data by Anterola and Lewis (77) determined, however, that the latter data accounted for $\geq 115\%$ of the thioglycolic acid lignin perhaps indicating that either the true lignin contents were grossly underestimated, or the alkaline NBO analyses were incorrect. This is an important matter as it again underscores the serious discrepancies in which many lignin “analyses” are often being carried out in this field. For example, currently many researchers typically do not either compare or contrast results between the different methods used for consistency, accuracy and/or quantification reliability. Accordingly, meaningful trends are potentially being missed in terms of overall effects of manipulation of lignin contents and compositions. Hence, taken together, the effects of *Arabidopsis* 4CL downregulation provided little insight yet other than perhaps (and quite unexceptionally) possibly reducing lignin levels. Again, these studies simply represent very preliminary analyses and now need to be extended to probe both effects on lignification proper and on vascular apparatus assembly.

4CL downregulation in aspen resulted in quite unusual assertions based on the data actually obtained (224). In that study, it was reported that the downregulated lines grew $\sim 50\%$ taller and that cellulose synthesis was markedly increased. Our reanalysis (77) gave a very different interpretation. In terms of accelerated growth, the lignin contents of the controls were 21.62%, whereas the “growth acceleration” reportedly occurred in the various transgenic lines having between 20.6 and 11.8% lignin, respectively (224). That is, a very small

reduction in lignin content from 21.62 to 20.6% was considered sufficient to result in accelerated growth. This effect has not been observed in any other study of lignin downregulation/mutation, and requires scientific explanation, if correct. Moreover, to date, there have been no further reports of effects on poplar growth by 4CL downregulation. These investigators, or others, should thus clarify at the earliest opportunity whether growth enhancements are still observed for both greenhouse and field-grown 4CL downregulated lines, and what effects on lignification, cell wall architecture, and physiology in general result. One possibility to be considered is that this plant line is capable of overproducing reaction (tension wood) tissue as a mechanism of partially “compensating” for reduced lignin levels, i.e., as noted for *pC3H* downregulation in alfalfa (72).

If so, this could further explain the reportedly enhanced cellulose contents, which had also only “increased” proportionally in large part because of reductions in lignin levels. Furthermore, although the full NMR spectroscopic data was not provided for that study (224), it was stated that the lignin in the 4CL downregulated line (reduced by ~45%) was apparently the same as that of wild type, as were also the G/S ratios by thioacidolysis. This would again provisionally be expected from first principles (77): namely that there were simply reductions in lignin content when 4CL activity was greatly suppressed without any “compensation” by other non-monolignol phenolic entities into the lignin core structure as suggested by others (173–175, 226). Finally, these preliminary studies again emphasize the need to conduct more extensive analyses of the lignins and to determine the effects on the vascular apparatus integrity in 4CL downregulated lines.

7.6.1.5 Summary

The effects of downregulating and/or mutating PAL, C4H, *pC3H*, HCT, and 4CL, as far as the monolignol/lignin pathway was concerned, gave predictable results. PAL, C4H, and 4CL resulted in lower lignin levels overall, with generally deleterious effects (weakening) of the vasculature being noted, albeit not quantified in any way. *pC3H*, a regulatory branch point to the G/S segments of the lignin-forming pathway, also gave quite predictable results, i.e., significant reductions (up to 64–68% of estimated wild-type lignin levels and/or altered vascular anatomy). Interestingly, for alfalfa, reaction (tension) wood tissue provisionally appeared to be part of a compensatory mechanism to help offset reductions in vascular integrity due to lower lignin contents. No evidence was obtained, though, for “combinatorial” biochemistry being in effect to any measurable or significant extent in any of these manipulations/mutations. That is, there was no evidence of any shift to formation of non-monolignol phenolic moieties to compensate for reductions in overall lignin amounts. There was, however, a relatively small increase in the level of H-lignin being formed which is considered due to limited substrate degeneracy during a proposed template-assisted polymerization, but only for the *pC3H* line.

7.6.2 CCR, CAD, F5H, and COMT downregulation/mutation, and the enigma of monolignol radical generation

Since the early 1930s or so, there have been a number of reports indicating that various agronomically important plant species, i.e., maize (*Zea mays*), sorghum (*Sorghum bicolor*),

and pearl millet (*Pennisetum glaucum*), can produce so-called brown-midrib mutants. The maize mutants (*bm1-bm4*) were spontaneous mutations (58–60) and were later shown to have lower lignin contents than wild-type lines (62), whereas both sorghum (227) and pearl millet (228) mutants were generated via chemical mutagenesis. For sorghum, out of the 19 individual mutant lines obtained, two (*bmr12* and *bmr18*) had estimated lignin contents reduced by 42 and 45%, respectively, when compared to wild type (227). Since then a spontaneous mutation has also been identified in sorghum (*bmr26*) (229). To our knowledge, none of these mutants have yet found commercial application since their discovery, due to defects, such as increased brittleness, increased lodging, delayed growth, and flowering (230).

Some of these lines have been characterized as mutations in CAD (*bm1*) (67) and COMT [*bm3* (66, 231) and *bmr12/bmr18/bmr26* (229)]. As noted earlier, our previous metabolic flux analyses (34) had suggested that under the conditions employed none of these steps would “normally” have a rate-limiting role in terms of carbon allocation to the monolignol/lignin pathway. Nor would modulation of F5H be anticipated to alter carbon (metabolic flux) allocation. Effects of manipulating each of these steps are thus discussed below.

7.6.2.1 CCR: tyramine derivatives are not chemical signatures of CCR downregulation/mutation, and abnormal lignins are not formed

CCR has recently been both downregulated and mutated in tobacco (*Nicotiana tabacum*) (232) and *Arabidopsis* (233), respectively. The phenotypes obtained for each line were, at first glance, quite similar: both were considerably dwarfed (131, 232) (Figures 7.13C and 7.13D). Such effects, as already noted earlier, are not always a typical consequence of downregulating lignin amounts and/or compositions. Nevertheless, by comprehensively examining the lignin contents and compositions, stem diameters/lengths and anatomy, it was considered that the *Arabidopsis irx4* mutant was delayed in overall development (Figure 7.12F), particularly as regards lignification (131). Specifically, the deposition of S-lignin components in the *irx4* mutant line initially lagged behind that of wild type. At maturation, the lignin S/G compositions (ratios) were similar though for both lines, with the overall lignin amounts only being 10–15% lower in the mutant (131). On the other hand, the lignin levels reported for the CCR downregulated tobacco stems were *circa* 50% of wild-type levels, with the xylem cells collapsed (Figure 7.12E), indicative of a much compromised vascular apparatus (77, 232). At the anatomical level, it was of interest that both the *Arabidopsis* and tobacco stem cross-sections apparently differed substantially in their overall effects on vascular integrity; in neither case were “perfectly viable” plants obtained, given the defects/pleiotropic effects noted.

Detailed analysis of lignin deposition in the *Arabidopsis irx4* line had thus established that initially the mutant had a delayed but coherent (normal) program of lignification (131, 132). By contrast, a previous study by other researchers (233) had reported that this plant line contained an “abnormal lignin,” derived from “alternative” phenolics and whose lignin levels were reduced by *circa* 50%. The “abnormal” nature of the lignin was apparently determined by NMR spectroscopic analyses, although no data were actually provided. These reports have since been revised with more in-depth analyses from our laboratory (131, 132). First, the full extent of the lignification response in the *irx4* mutant line was only

actually determined by examining the lignin contents and compositions up to maturation (>8 weeks), with only a small reduction (~10–15%) in deposition levels noted. Second, NMR spectroscopic analyses indicated that typical G-S lignins were being formed and not “abnormal” lignins as had been reported. Furthermore, we proposed that the dwarfing phenomenon and reduced lignin levels may be due to CoASH levels being reduced (due to build up of hydroxycinnamoyl CoA derivatives) in the *irx4* line, with some sort of, for example, feedback inhibition occurring (131); however, this remains to be fully established in future studies.

Perhaps most importantly, plotting lignin contents versus (thioacidolysis) releasable monomers (representing a subset of cleavable 8–O–4'-linkages) again indicated that both *irx4* and wild-type lines had a monomer invariant frequency (Figure 7.14C) of said linkages at all stages of lignin deposition, plant growth, and development. These data are thus again discussed later in terms of further indications of a non-random assembly process.

The original, albeit now incorrect, report by Jones *et al.* (233) was apparently based on expectations that had been raised from the study of the CCR downregulation in tobacco (173, 174, 226). The latter papers described bewildering findings as regards lignin macromolecular assembly and composition. Specifically, it was reported that when CCR was downregulated in tobacco, the plants compensated for reduction in monolignol (lignin precursor) supply by incorporating other “alternative” phenolics into the lignifying cell walls. Initially, the “non-traditional” phenolics reported as incorporated into lignin through CCR downregulation included ferulic (11) and sinapic (13) acids, as well as a variety of other phenolics, including acetosyringone (61) (173, 174, 226). It was also reported that feruloyl tyramine (60) was “heavily incorporated” into the lignin as a consequence of CCR downregulation (173, 174, 226), and that this moiety represented a chemical “signature” for CCR downregulation (174). However, no quantification of feruloyl tyramine (60) levels of any sort was carried out.

As regards the reported increases in amounts of hydroxycinnamic acids/benzaldehydes, etc. in the lignins of the highly dwarfed, lignin reduced (~50%), CCR downregulated tobacco lines, relative to the wild type, there was apparently no significant difference in total amounts measured quantitatively (77). For example, the reported levels of such moieties ranged from 0.04 to 0.07% of stem cell wall residues in both wild type and downregulated lines (77). Such minute levels would not constitute compelling evidence for “abnormal” lignin and “aberrant” lignins being formed from “non-traditional monomers.” Indeed, since the lignin contents of CCR downregulated and wild type were ~11 and ~22% of the plant stem dry weight, respectively (77), the amounts of the aldehydes/acids (0.04–0.07%), etc. were minuscule relative to actual lignin levels.

Other studies by Ralph and colleagues (173, 174), which reported that there were elevated levels of feruloyl tyramine (60) covalently attached to tobacco lignins, and that these were chemical signatures of CCR downregulation/mutation, could not be independently verified either. Furthermore, the lack of any feruloyl (*p*-hydroxycinnamoyl) tyramine (60) resonances in the *Arabidopsis* CCR-*irx4* lignin-derived isolates (77, 132) eliminated these as being generic chemical “signatures” for CCR mutation/downregulation as had been proposed (174). By contrast, careful isolation (by dissection) of the vascular (lignified) apparatus and subsequent detailed NMR spectroscopic analyses of the resulting lignin isolate(s) gave no evidence for the presence of feruloyl tyramine (60) moieties in the lignins from the wild-type lines of tobacco (177). Nor were feruloyl tyramine moieties (60) observed in the lignin

preparations by Bernard-Vailhé *et al.* (234). Additionally, in contrast to reports (173, 174) of feruloyl tyramine (**60**) moieties being incorporated into lignin, these researchers had never demonstrated that the feruloyl tyramine-like resonances in the lignin-enriched isolate were covalently linked to lignin, or came from the same cell wall types harboring lignins. Nor was evidence provided that the overall amounts of feruloyl tyramine (**60**) moieties had increased in the isolates from the CCR downregulated lines, relative to the original levels in the wild-type line.

7.6.2.2 CAD

Catalyzing the final step in monolignol 1–5 biosynthesis, the effects of downregulation and mutating *bona fide* CAD genes have been extensively studied and reported upon [see Anterola and Lewis (77), for a comprehensive discussion and analysis]. The common phenotype resulting from CAD mutation/downregulation is that of a red-brown coloration in the xylem region (57, 71, 172, 235–237) and Figure 7.12G shows this effect in a stem cross-section from a double CAD mutant (*cad-4 cad-5*) in *Arabidopsis* (57, 71). As mentioned above, this coloration has been known for nearly eight decades (58) in the brown midrib mutants, with the *bm1* associated with expression of the CAD gene (67). This red-brown coloration was also reported as being due to formation of an abnormal “wine-red” lignin in tobacco (236, 237), with this being proposed to have good potential for furniture staining, dyes, and so forth. Moreover, the results from such studies and other analyses of presumably lignin-enriched isolates from CAD downregulation/mutation of tobacco were rationalized by several investigators as further evidence for random coupling/combinatorial biochemistry, and thus of lignin’s composition not being particularly important (173, 226).

Our own data and the interpretations thereof, provide very different findings and insights. More importantly, they give an evolutionary perspective as to why lignins are essentially monolignol (**1**, **3**, and **5**) derived (and, partially from monolignol esters **30–32** in grasses as well). Specifically, the recent findings now help explain why lignins are not formed from, for example, *p*-hydroxycinnamaldehydes **19**, **21**, and **23**.

7.6.2.2.1 “RED LIGNIN”: A MISNOMER

Comprehensive analyses and reassessment of reports of “red lignin” in tobacco, (235–237) resulting from CAD downregulation, established that there was no “red-lignin” as such (177). Instead, it was simply a pigment [mainly sinapyl aldehyde (**23**) derived] in near trace amounts in tobacco that could readily be removed under conditions generally used for floral pigment removal (i.e., 0.5% HCl in MeOH) (177). Similar treatment of *Arabidopsis* “red xylem” also resulted in facile removal of this coloration, with concomitant release of sinapyl aldehyde (**23**) (71). No evidence was obtained, though, this red pigment was an integral part of the polymeric lignin. Interestingly, the red coloration could also be reconstituted on either preextracted tobacco xylem tissue cross-sections and/or polyamine TLC plates, by dipping either into a dilute solution of sinapyl aldehyde (**23**) (177). Furthermore, Bernard-Vailhé *et al.* (234), using modified Björkman procedures to isolate lignin derivatives from both wild type and CAD downregulated tobacco xylem, had also noted that the red coloration was completely removed from the lignin-derived preparations. These data also contrast with a contribution by Boerjan and coworkers (238) who reported that the “red xylem” coloration in tobacco could not be removed by treatment with sulfuric acid, methanol,

butanol/HCl, acetyl bromide, or triethylene glycol. Such an observation, therefore, needs to be independently confirmed, as it appears incongruous with the findings using both tobacco and *Arabidopsis*.

7.6.2.2.2 CAD DOWNREGULATION/MUTATION IN LOBLOLLY PINE, TOBACCO AND ARABIDOPSIS: PHYSIOLOGICAL EFFECTS AND NMR SPECTROSCOPIC ANALYSES

In the study of effects of downregulating/mutating CAD in loblolly pine and tobacco, there have been various reports of “abnormal” lignins being produced. These have been described as affording lignin-enriched isolates harboring significant levels of incorporated hydroxycinnamaldehydes **19**, **21**, and **23** and hydroxybenzaldehydes **45–47**, albeit without quantification to indicate what “significant” meant (173, 239).

7.6.2.2.3 CAD MUTANT LOBLOLLY PINE, 2-METHOXYBENZALDEHYDE AND DIHYDROMONOLIGNOL/LIGNIN DERIVATIVES

One of the unusual findings apparently made initially as regards a CAD mutant loblolly pine isolate was that it biosynthesized 2-methoxybenzaldehyde (**62**, Figure 7.11A) (172); this was a quite unexpected finding given that such a metabolic response would not have been anticipated, from first principles, to result from CAD downregulation. This finding was then retracted (173), being a consequence of an error in the NMR spectroscopic assignments. Additionally, various dihydroconiferyl alcohol (**63**, Figure 7.11A) containing substructures were reported as apparently present in the “abnormal” lignin (discussed below), linked to a “conventional lignin monomer,” thereby indicating that at least 50% of the monomers were still monolignol-derived (172). Such moieties with a reduced side-chain had, however, been previously described in *Pinus taeda* (loblolly pine) cell suspension culture (171). Indeed, the allylic double bond reductases responsible for this reduction step, and the encoding genes have since been fully reported by our laboratory from both *P. taeda* (240, 241) and *Arabidopsis* (242); the X-ray structure of the ternary complex has also been obtained and the proposed biochemical catalytic mechanism described (242). Moreover, the CAD loblolly pine mutant has been described as presumably harboring a number of mutations (74), with the plant line apparently stunted. There are no reports of effects on the vascular integrity of this loblolly pine CAD mutant, which is presumed to be weakened, assuming that lignification proper has been modulated/reduced.

7.6.2.2.4 CAD DOWNREGULATED TOBACCO: NMR SPECTROSCOPIC ANALYSIS AND PRELIMINARY EFFECTS ON VASCULAR INTEGRITY

CAD downregulated tobacco lines were preliminarily studied by several researchers (235–237, 243), who concluded that overall gross lignin contents were unaffected by CAD downregulation. As discussed in Anterola and Lewis (77), there were significant differences though in the presumed overall polymeric architecture/properties of the lignified tissue, as evidenced by increased susceptibility to alkaline treatment (77, 243). Additionally, various CAD downregulated lignin-enriched isolates were subjected to detailed NMR spectroscopic analyses (173, 177, 234). In our own analyses, the extent of the aldehyde-enhanced resonances present in the CAD downregulated tobacco [generated by Halpin *et al.* (235)] was very modest (177). Quantification, as determined by NMR spectroscopy (estimated from the normalized ¹H NMR spectra with respect to the aromatic region of lignin isolates), gave small amounts of *p*-hydroxybenzaldehydes **45–47**, *p*-hydroxycinnamaldehydes **19/21/23** and 8–O–4′-linked

p-hydroxycinnamaldehyde (substructure **XIIa–c**, Figure 7.11B) end groups in the lignin isolates: ~0.68, 1.5, and 1.41% of total lignin, respectively (177).

In terms of overall effects on the vascular integrity, preliminary studies have also yielded quite predictable results, in the sense of weakening of the vascular apparatus (68–70, 244). In particular, the antisense CAD line had reduced longitudinal tensile stiffness and lower shear and Young's moduli, respectively, compared to the controls (68). Moreover, these lines apparently responded to periodic flexural stem bending by increasing the thickness of the xylem cylinder (70), although none of the anatomical data (e.g., cross sections and cell types affected) were described to give insight into the changes occurring. Incongruously, as described earlier, Hepworth *et al.* (69) reported that the downregulated tobacco lines had lignin contents of 40–50%, again further reflecting the ongoing problems experienced in this field for reliable measurement of lignin contents/composition: generally, for tobacco, lignin contents are in the range of ~20–25% at maturation. Nevertheless, once again the effect of manipulation of a gene involved in the monolignol/lignin pathway had deleteriously altered structural properties of the lignocellulosic matrix.

7.6.2.2.5 CAD DOUBLE MUTATION IN ARABIDOPSIS: COMPREHENSIVE ANALYSES OF PHENOLIC CONSTITUENTS AND EFFECTS ON VASCULAR APPARATUS INTEGRITY

A *cad-4 cad-5* (*cad-c cad-d*) double mutant of *Arabidopsis* (ecotype Wassilewskja) was also recently successfully generated (57) and subsequently comprehensively analyzed (71). Figure 7.13E shows the prostrate phenotype that results from the double mutation relative to that of wild type (71); as anticipated, the dynamic modulus properties were also substantially reduced, this being a further indication of a structurally weakened vascular apparatus (71). Several other important features were identified following the comprehensive study of different stages of growth/development and polyphenolic deposition into the cell walls of the double mutant, using both chemical degradation and ¹³C NMR spectroscopic analyses. Specifically, the double mutant contained very small amounts of monolignols (*circa* 10% of the polyphenolics present), as well as polymeric *p*-hydroxycinnamaldehyde moieties. At plant maturity, these constituted together about 11.3% of the cell wall residue, in contrast to lignin in the wild-type line which was almost double this amount (~22.5%). Chemical degradation (thioacidolysis) analyses though established that the total amount of monomer-cleavable 8–O–4' interunit linkages, i.e., monolignol **1**, **3**, and **5** and styryl–O–aryl ether (substructure **XII**, Figure 7.11B), etc. derived substructures closely mirrored that for monolignol-derived lignin deposition (Figure 7.14D) at the same stages of growth/development. [This was not previously recognized by other researchers (57) since their analyses of the styryl–O–aryl ether-derived substructures (**XII**, Figure 7.11B) were lower by almost an order of magnitude, due to a lack of either authentic standards and/or correct response factors.]

That is, the 8–O–4' interunit linkage frequency was again apparently directly proportional to polyphenolic content in a manner somewhat analogous to that of lignification proper in the wild-type line; however, the *p*-hydroxycinnamaldehyde deposition was terminated/aborted prematurely (see arrowhead, Figure 7.14D). Perhaps significantly, this metabolic “checkpoint” – arresting polyphenolic deposition – also appears to be coincident to that of termination of H-lignin deposition in the *pC3H* line, suggesting a common mechanism is in place for terminating/aborting both processes. Interestingly, 8–5' linkages were also very evident in the *p*-hydroxycinnamaldehyde isolates, but in this case with substructure **IX** (Figure 7.11B); others [e.g., substructures **X** and **XI** (239)] were not detected (71).

Taken together, it is proposed that the double mutant has attempted, in a futile manner, to produce a *poly-p*-hydroxycinnamaldehyde facsimile (of inferior structural properties) to that of monolignol 1-, 3-, and 5-derived lignins. This can thus be *provisionally* envisaged to occur through very limited substrate degeneracy during template polymerization, that would normally be operative for constitutive macromolecular lignin assembly. On the other hand, the poor biophysical/structural integrity properties of the phenotype so obtained presumably provide useful insight as to why *poly-p*-hydroxycinnamaldehydes did not evolve as a substitute for the monolignols 1, 3, and 5 in lignification proper.

7.6.2.2.6 “CAD1,” A MIS-ANNOTATION?

Downregulation of the currently annotated “CAD1” homologues, *NtCAD1-1* and *NtCAD1-7*, in tobacco using an RNAi technology (141) recently led to selection of two lines, L11 and L14: expression of the “*NtCAD1*” genes was barely detectable in line L11, whereas it was reduced in line L14. Lignin contents, as estimated by the Klason method, were though basically unchanged in both lines as compared to wild type. While these researchers proposed that the G contents in the lignin were reduced in the transformants, the body of evidence was scant for any effect on lignification proper. Indeed, the observations made may yet again be another consequence of pleiotropic effects. These data are in marked contrast to the striking effects of “knocking-out” both *AtCAD4/5*, the corresponding *bona fide* CADs present in *Arabidopsis*, the result of which was that neither the corresponding monolignols nor lignin were formed to any extent (71). Thus, the involvement of the putative tobacco “CAD1s” in both coniferyl alcohol (3) and lignification has not been established; other physiological roles need to be considered and examined as indicated earlier.

In this respect, a recent study of two genes (*LePAR1* and *LePAR2*) from tomato (*Solanum lycopersicum*), with high degree of similarity/identity to both the *Eucalyptus* and tobacco CAD1s resulted in a different metabolic role being identified: Tieman *et al.* (245) concluded that these genes encoded a 2-phenylacetaldehyde reductase required for formation of 2-phenylethanol (64, Figure 7.11A). Interestingly, both *Eucalyptus* and tobacco accumulate 2-phenylacetaldehyde (65)/2-phenylethanol (64) as volatile constituents; these data thus suggest that their CAD1s may instead be 2-phenylacetaldehyde reductases.

7.6.2.3 F5H: effects of substrate degeneracy during proposed template polymerization

Ferulate 5-hydroxylase (F5H) is not considered to have a rate-limiting role in carbon allocation to the monolignol/lignin-forming pathway in angiosperms (77). In agreement with this, a *fah1-2* mutant isolated in *Arabidopsis* from an ethane methyl sulfonate (EMS) mutagenized population (246) had a mutation in the F5H gene (116) and gave a plant line with apparently a similar level of a guaiacyl-rich lignin rather than the guaiacyl-syringyl lignin of the wild type (247). In this laboratory, it was thus again considered instructive to compare the amounts of monomers/dimers released by thioacidolysis as a function of lignin deposition at different stages of “stem” growth and development until maturation (188). The results obtained are summarized in Figures 7.14B and 7.14E–J. Again a linear relationship between lignin contents at each stage of growth/development (of the stem sections) and of the amounts of monomer/dimer-releasable 8–O–4', 8–1', 8–5', 5–5', and 5–O–4' interunit

moieties for both wild type and *fah1-2* lines were observed. That is, for the 8–O–4' interunit linkages, at least as far as releasable monomers are concerned, there was a linear correlation between their frequencies and lignin contents, regardless of whether wild type or *fah1-2* lines (Figure 7.14B). A similar relationship also held for 8–1', 8–5', 5–5', and 5–O–4' interunit linkages (quantified as dimers). By contrast, no significant levels of pinoresinol-like (8–8') substructures (**IIIb**, Figure 7.2D) were detectable in the *fah1-2* line following thioacidolysis (Figure 7.14J), whereas the wild-type line resulted in the release of relatively small amounts of syringaresinol-derived substructures (**IIIc**, Figure 7.2D). [This observation was again in keeping with the well-known “conundrum” that pinoresinol-like substructures (**IIIb**, Figure 7.2D) are presumably covalently modified in some way thereby preventing their release during this treatment. However, as indicated earlier, this again contrasts with synthetic dehydropolymerizates (DHPs) which readily release these substructures; this is considered as indicative of further differences from lignin formation *in vivo*.] It should be emphasized though that *in toto* our current best estimates of the percentage of the total interunit linkages in wild-type lignins being accounted through monomer/dimer release are only *circa* 43%: 8–O–4' (27.1%); 8–1' (4.5%); 8–5' (5.4%); 5–5' (4.9%); 5–O–4' (0.9%); and 8–8' (0.3%, S only), whereas in the *fah1-2* they are *circa* 44%: 8–O–4' (28.1%); 8–1' (4.1%); 8–5' (5.4%); 5–5' (5.1%); 5–O–4' (1.19%); and 8–8' (not detected).

The *fah1-2* mutant was also genetically modified under control of either the cauliflower mosaic virus (CaMV) 35S or *A. thaliana* *C4H* promoters, respectively, the purpose of which was to attempt to generate a syringyl-rich lignin (247); of these, only the C4H-F5H line afforded a significantly increased S-enriched tissue. While the syringyl contents of the lignifying stem tissues were clearly increased, lignin contents and monomeric analyses (NBO) were inexplicably determined at different time intervals, i.e., at 8 and 5 weeks, respectively (247, 248). This is again a further indication of the difficulties experienced in this field as regards lignin analyses, and the urgent general need for both chemical reliability and more incisive testing protocols.

It is important to note that overexpression of the *F5H* gene (with the *C4H* promoter) would be expected to occur in all cell types involving metabolism of *p*-coumaric acid (**9**), e.g., whether to lignins, lignans, hydroxycinnamic acids, suberins, flavonoids, etc. For the monolignol-forming pathway, its overexpression would be expected to provide (in a G- and/or G/S lignin-forming cell), the monolignol 5-hydroxyconiferyl alcohol (**4**) as discussed earlier (77). The latter could then either be deployed as such in lignification and/or converted into sinapyl alcohol (**5**) – if there was an OMT expressed concurrently able to methylate 5-hydroxyconiferyl alcohol (**4**) [see Anterola and Lewis (77) for a previous discussion]. However, as comprehensively discussed in Anterola and Lewis (77), the participation of 5-hydroxyconiferyl alcohol (**4**) in ordered lignin assembly would not be expected to fully duplicate that of sinapyl alcohol (**5**), i.e., the catechol nature of 5-hydroxyconiferyl alcohol (**4**) would presumably change the *chemistry of any programmed template polymerization*. This is because the 5-hydroxyconiferyl alcohol (**4**) would be anticipated to be more susceptible to quinone formation and/or, alternatively, to alter the outcome of free-radical coupling, e.g., to generate benzodioxane-like substructures such as substructure **XIV** (Figure 7.11B). Yet, based on the initial Marita *et al.* study (248), this did not happen. The syringyl content of the lignin isolates was initially considered to be >90%, with <5% guaiacyl-derived moieties being detected (determination method not described); 5-hydroxyconiferyl alcohol (**4**)-derived moieties were not reported as present. This report was then modified later to

indicate that 5-hydroxyconiferyl alcohol (**4**) moieties were present in the lignin to ~10% as determined by estimation of NMR contour volumes (249); this change again reflects both methodology and analysis limitations.

In this regard, the data obtained upon comparison of lignin contents, at various stages of growth/development, versus thioacidolysis releasable monomers/dimers, are again summarized in Figures 7.14B, E, F, and J for the C4H-F5H *Arabidopsis* line. Note that at maturity, the estimated H:G:S ratios by thioacidolysis were 1:7:47. Although methodology is currently being developed to precisely quantify to what level 5-hydroxyconiferyl alcohol (**4**)-derived moieties are present in the resulting lignin(s), the amounts are *provisionally* considered to be <5–10%. Taken together, the frequencies of 8–O–4' interunit linkage-cleavable monomers and lignin contents in the C4H-F5H transformant are thus quite comparable to that of wild type and *fah1-2*, i.e., again indicative of polymer formation leaning toward monomer-invariant linkage frequency. The same appears to be the case for the 8–1' linkages (quantified as releasable dimers). Interestingly, the syringaresinol substructures (**IIIc**, Figure 7.2D) were the only other S-units fully identified as releasable dimeric products thus far (Figure 7.14J).

All of the 8–O–4' (26%), 8–1' (4.5%), and 8–8' (1.5%) of monomer/dimer-cleavable linkages only account though for *circa* 32% of the lignin in the plant tissue (current best estimates), relative to ~43% in the wild-type line. Such low recoveries are not to be expected for a lignin presumed to contain essentially only S-moieties in 8–O–4', 8–1', and 8–8' interunit linkages, as has been proposed. Indeed, even correcting for *circa* 5–10% of 5-hydroxyconiferyl alcohol (**4**) amounts, the recoveries are still low. In short, the data for all of the transformants obtained have thus far only provided a very limited view of how macromolecular lignin assembly/configuration is controlled in the C4H-F5H line, i.e., whereby the recoveries are not at an acceptable level relative to the entire polymeric lignin present. Furthermore, the ability to modulate S/G levels presumably again reflects limited substrate degeneracy of the G and S monolignols and to a lesser extent 5-hydroxyconiferyl alcohol (**4**) during proposed template polymerization. However, what has exactly happened, as far as proposed template polymerization is concerned, remains to be elucidated through, for example, primary lignin structure determination.

7.6.2.4 COMT: further constraints to lignin macromolecular configuration

The maize *bm3* mutant (61), now known to affect COMT activity (66), is a plant line with reduced stem stalk strength and prone to stalk lodging (falling and collapsing) relative to wild type, thereby currently precluding its commercial utilization (250) [see discussion in Anterola and Lewis (77)]. Not having a rate-limiting function in the pathway in terms of carbon allocation, modulation of COMT activity does not affect overall lignin contents. Instead, the *monolignol*, 5-hydroxyconiferyl alcohol (**4**), is formed rather than sinapyl alcohol (**5**); the former is then used in place of sinapyl alcohol (**5**), but whereby a polymeric structure is generated whose properties are such that it results in a weakening of overall plant vascular integrity due to the incorporation of (**4**) into the lignifying matrix. This is again considered to reflect a tightly programmed lignification response, albeit where the outcome is affected by a limited substrate degeneracy during template polymerization. This, however, results in

diminished cell wall properties and helps explain why 5-hydroxyconiferyl alcohol (**4**) never evolved to be an effective lignin precursor.

As indicated earlier, the role of COMT in the second methylation step of syringyl lignin biosynthesis was first demonstrated by Atanassova *et al.* (136) in tobacco, and later further confirmed by others in poplar (182), aspen (251), and alfalfa (252). Downregulation/mutation of COMT though had no apparent adverse effect on overall lignin amounts as indicated using Klason lignin estimations (77), with the striking exceptions reported from the Dixon laboratory (186, 212). These latter researchers had again used the unreliable thioglycolic acid lignin (186) and the neutral detergent fiber (NDF)/Klason lignin methods (212), the results of which appeared to indicate that lignin contents had been (significantly) reduced (by *circa* 60%). While this interpretation can now be viewed as incorrect, this was the same methodology used for the PAL (207, 212) and C4H (207) studies; hence, the overall findings from these studies should also be reexamined with more robust and reliable technologies.

Additionally, in preliminary studies, COMT downregulation had no apparent significant adverse effect on the amounts of thioacidolysis-releasable G-monomers in the lignins from either tobacco or poplar, given that the total amounts, relative to lignin contents, were very similar to the wild-type lines (77). On the other hand, little to no S-units could be detected. To account for this, we can *provisionally* propose that in the COMT downregulated lines, the 5-hydroxyconiferyl alcohol (**4**) moieties were undergoing homo-coupling, whereas in the wild-type lines, sinapyl alcohol (**5**) was also doing the same. However, instead of 8–O–4' bond formation, etc. occurring as in the wild-type line (see Figure 7.15A), the presence of the flanking 5-hydroxyl group (from **4**) has resulted in an apparently near-quantitative replacement with the benzodioxane substructure (Figure 7.15B), rather than the “simple” 8–O–4' interunit linkage present in the wild-type line. That is, the main difference between the 8–O–4' linkage (Figure 7.15A) and the benzodioxane substructure formation (Figure 7.15B) is only in “trapping” of the quinone methide intermediate during template polymerization. [By contrast, if the incoming monolignol radical was a *p*-coumaryl (**1**), coniferyl (**3**), or sinapyl (**5**) alcohol moiety, the “trapping” would be performed by an external water or possibly carbohydrate molecule.] Nevertheless, formation of the benzodioxane substructure can thus still be considered as an 8–O–4' interunit linkage, even if the overall substructure generated differs.

Furthermore, with the *Arabidopsis* COMT-downregulated line, there *provisionally* appears to be a near equivalent reduction in amounts of G-releasable monomers to that of the original S levels (see Figure 7.15C which plots G and S 8–O–4' interunit cleavage, leading to monomer release, at different stages of *Arabidopsis* growth and development; Jourdes *et al.*, manuscript in preparation). In this case, the near *equivalent* reduction of both G and 5-OHG (previously S) moieties at different stages of growth and development suggests that G-S hetero-coupling was mainly occurring in the wild-type line, i.e., via hetero-coupling of coniferyl (**3**) and sinapyl (**5**) alcohol-derived monomers, thereby affording cleavable 8–O–4' interunit linkages. This is now apparently replaced by an equivalent level of hetero-coupling between coniferyl (**3**) and 5-hydroxyconiferyl (**4**) alcohol-derived radicals to afford the mixed benzodioxane substructure (Figure 7.15B). This near 1:1 reduction in releasable G:S moieties thus places another considerable constraint on how lignin macromolecular configuration is actually being achieved. It could, for example, possibly indicate that G and S monomers are being alternately laid down during macromolecular lignin assembly;

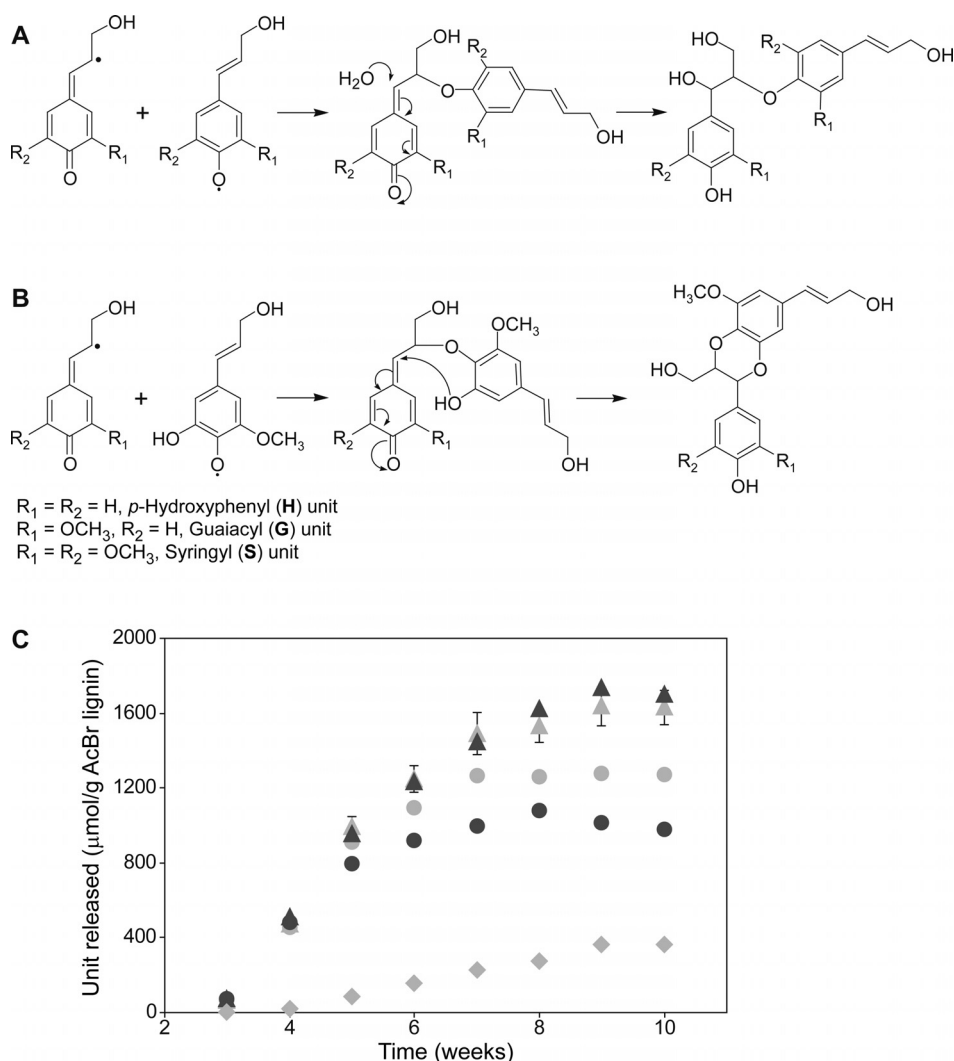


Figure 7.15 Formation of 8-*O*-4' interunit and benzodioxane linkages. (**A**) H, G, and S monolignol coupling and (**B**) H, G, 5-OH-G, and S monolignol coupling with 5-OH-G (**4**) to give benzodioxane (8-*O*-4') substructures. (**C**) Plots of thioacidolysis monomers H, G, and S (as $\mu\text{moles/g}$ AcBr lignin), released from both *Arabidopsis* wild type and COMT mutant lines, at different stages of growth and development. Grey \bullet , \blacklozenge , and \blacktriangle represent amounts of G, S, and G+S released monomers from wild type, whereas black \bullet represents releasable G-moieties in COMT mutant. Note that the differences between wild type and COMT levels reflect the loss of equivalent amounts of 5-OHG + S moieties; this is shown as a black \blacktriangle .

such observations once again underscore the urgent need for technology development to determine lignin primary structure(s).

In alfalfa, however, COMT downregulation resulted in reduction of both G-S and S-S amounts, this in turn presumably reflecting *differing cell-type* lignification processes. In this regard, although the incorporation level of 5-hydroxyconiferyl alcohol (**4**) moieties into the

lignin was not quantitatively estimated using chemical lignin degradative analysis, attempts to do this indirectly were carried out, i.e., by attempting to estimate the level of benzodioxane substructures using NMR spectroscopic analyses of the extracted lignin-enriched isolates by measuring volume integrals in the 2D HMQC spectra (253). Reassessment of this data herein also results in the provisional conclusion that overall levels of 8–O–4' interunit linkages were actually conserved in both wild-type and COMT-deficient alfalfa lines. That is, the wild-type line apparently had *circa* 81% of 8–O–4' linkages, whereas the COMT-deficient alfalfa had ~38% of benzodioxane and 44% of 8–O–4' linkages. Together, these represent ~82% of substructures with 8–O–4' linkages, and thus represent a high degree of conservation of the overall 8–O–4' linkage type in both wild type and COMT lines.

These data are thus also in apparent harmony with a proposed limited template degeneracy in assembly of the lignin macromolecule – but where both the chemistry on the proposed template is altered in the mutation (to involve catechol, rather than phenolic, chemistry). It is worthwhile reflecting that the alfalfa data has been known since 2003 (253), and yet has been interpreted by others as evidence for random coupling/combinatorial chemistry leading to 10⁶⁶ isomers, etc. (175). On the other hand, while the COMT mutation results – in a presumably ordered way – in formation of benzodioxane substructures (Figure 7.15B), the biophysical effects, nevertheless, resulted in weakening of the overall vasculature.

7.6.2.5 *The enigma of monolignol radical generation*

A key factor in ordered lignin macromolecular assembly/configuration is in temporal and spatial control over monolignol radical generation (discussed below as regards template-facilitated polymerization). Since the 1950s, various oxidases have been implicated as having roles in lignification – these have included peroxidase (12, 53, 54, 171, 254–257), laccase (12, 53, 54, 258–266), combined peroxidase and laccase (12, 53, 54, 254, 267), coniferyl alcohol oxidase (268–271), (poly)phenol oxidase (272–278), and cytochrome oxidase (279). Their putative involvement has generally relied upon the ability of such enzymes to oxidize monolignols in vitro, even though none of these enzymes have yet engendered formation of products in vitro that duplicate faithful facsimiles of lignin structure; in some cases, they do not even generate formation of polymeric products. [The reader is again encouraged to review the historical developments as regards consideration of each of these oxidases and their potential for lignification and the mechanistic questions that they raise (31)].

7.6.2.5.1 *PROTEINS WITH NO YET ESTABLISHED ROLES IN LIGNIFICATION: POLYPHENOL OXIDASES, CONIFERYL ALCOHOL OXIDASE, AND LACCASES*

None of these three enzyme classes have any demonstrable role in lignification. The early suggestions that a polyphenol oxidase might be involved emerged from studies by Freudenberg in 1953, using a press–sap extract of the mushroom, *Agaricus campestris* (272–277) and also later by Mason and Cronyn (278); mushrooms do not, however, biosynthesize lignins. Another candidate was coniferyl alcohol oxidase, detected in jack pine (*Pinus strobus*) (268, 269), other *Pinaceae* species (268, 269, 271) and tobacco (270); this has also not been demonstrated to either afford lignins in vitro and/or have a role in lignification in vivo.

As of 2007, there was still no convincing evidence for any direct involvement of laccase in lignification – in spite of numerous articles (12, 53, 54, 258–266) appearing over a five

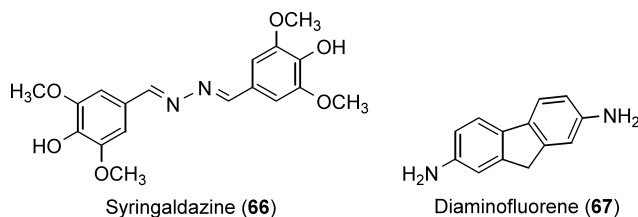


Figure 7.16 Chromophoric substrates.

decade plus time-frame supporting their involvement in lignin macromolecular assembly. None of these studies, however, met the criteria for monolignol oxidation/lignification as set out by Lewis *et al.* (31), namely, that: “the enzyme must be able to convert mono-, oligo- and polyphenols into their free-radical derivatives; the enzyme must be both temporally and spatially correlated with sites of lignin biosynthesis; the enzyme, in the presence of the requisite co-factors and/or other proteins, must be demonstrably capable of converting the monolignols into macromolecular lignin chains; and the enzyme must unequivocally be demonstrated as essential for lignin biosynthesis, e.g., through loss of function.”

The evidence for a role of laccases in lignification was scant indeed: originally proposed by Russian workers in the 1940s (258, 259), this was later investigated further in the laboratories of Freudenberg (12, 53, 54) and Higuchi (260). While laccases are generally capable of oxidizing monolignols, the experiments were not designed at that time to establish if native lignins were being formed with these catalysts. Later, Harkin and Obst (255) reported the exclusive participation of peroxidases in lignification of trees, using the reagent syringaldazine (66, Figure 7.16). They also concluded that laccases, for example, were absent in *Acer* (sycamore) species examined based on a lack of histochemical staining. Laccases were, however, subsequently purified from *Acer pseudoplatanus* (261, 263, 267), and the encoding genes cloned (265), thereby disproving their reported absence. On the other hand, as far as putative lignification was concerned, incubation of monolignols with the *Acer* laccase(s) (263, 267) *in vitro* only gave preparations with very minor amounts of 8–O–4' interunit linkages that were presumably not polymeric; that is, such preparations did not reflect lignin structure which predominates in 8–O–4' interunit linkages.

An additional study (264), purportedly detecting a laccase in loblolly pine suggested that laccases were also associated with lignification on the basis of staining with syringaldazine (66) and diaminofluorene (67). Furthermore, reaction rates (as measured by rates of oxygen consumption) for *p*-coumaryl (1), coniferyl (3), and sinapyl (5) alcohols were 5, 72, and 47 nkat mg⁻¹ protein, respectively, with *K_m* values for each either being unobtainable as for *p*-coumaryl alcohol (1), or very high 12 and 25 mM for coniferyl (3), and sinapyl (5) alcohols, respectively. Such data prompted Ros Barceló (280) to comment “With these high *K_m* values, it is difficult to imagine what concentration of cinnamyl alcohols it would be necessary to reach its lignifying cell walls to saturate laccase during the oxidation of cinnamyl alcohols to lignin-like compounds.” These data do not therefore represent proof of laccase involvement in lignification.

Later studies (281), attempting to downregulate the laccase multigene family in poplar (*Populus trichocarpa*), had essentially no effect on lignin contents as estimated by both

acetyl bromide and Klason methods (20–25% of CWR), or on lignin compositions as determined by thioacidolysis. These findings again indicated that laccases have no significant and/or direct role in monolignol 1/3/5 oxidation leading to macromolecular lignin assembly/configuration, in contrast to more than five decades of scientific contributions (12, 53, 54, 258–266) proposing the contrary view.

Interestingly, *Arabidopsis* has 17 genes encoding laccases, and each of these has been examined for patterns of gene expression using the GUS-reporter system as before (Turlapati *et al.*, manuscript in finalization): eight of these are expressed in vascular (lignifying) tissue(s), although their physiological roles still need to be defined. One of the laccases apparently has a role in seed coat development, with this presumed to be required for condensed tannin formation (282). Yet, at the time of writing, other researchers still continue to suggest that laccases have a role in seed coat lignification (266), even though seed coat tissue apparently does not form lignin.

7.6.2.5.2 PROTEINS WITH ESTABLISHED ROLES IN OXIDATION/POLYMERIZATION: PEROXIDASES

These are large multigene families in plants whose full physiological/biochemical roles and functions still remain poorly understood. Unlike laccases, however, a direct role for peroxidases in lignification has been demonstrated. That is, downregulation of a peroxidase (TP60) in tobacco (*N. tabacum*) gave transformants with lignin levels reduced by *circa* 40–50% (257). Additionally, phloroglucinol-HCl staining suggested that the vasculature had been weakened (Figure 7.12H), although no quantitative structural testing on the plant stems was carried out. Thus, at present, the only oxidative enzyme demonstrated to have a role in monolignol oxidation/lignification is peroxidase. As before, more detailed analyses are required to fully ascertain the effects of peroxidase downregulation on lignification/cell wall structure(s).

7.6.2.6 Summary

Studies of CCR downregulation/mutation gave rise to severely dwarfed phenotypes; in *Arabidopsis*, typical G/S lignins were biosynthesized albeit at a delayed rate. There was no convincing evidence for replacement of monolignols with other non-monolignol moieties – e.g., feruloyl tyramine (**60**), acetosyringone (**61**), etc. The studies with CAD, F5H, and COMT were also most informative: while many of the mutants (e.g., CAD, COMT) have been known for almost three-quarters of a century, the biochemical basis of how they (CAD and COMT) disrupt the normal proposed template polymerization (see later) has now apparently come to light, i.e., via limited substrate degeneracy on the proposed lignin-forming template. The effects of this (attempted) degeneracy were not though structurally beneficial, and thus help explain why neither *p*-hydroxycinnamaldehydes nor 5-hydroxyconiferyl alcohol (**4**) evolved as substrates proper for lignification. Additionally, the F5H mutant (*fah1-2*) and the C4H-F5H overexpressing lines afforded two lignins with altered G and S levels, as did downregulation/mutation of COMT. Significantly, the patterns of interunit linkage frequency established that (based on monomer/dimer release) lignification was apparently proceeding in similar (controlled) manner in each case. The data are explained through limited (substrate) degeneracy during proposed template polymerization.

7.6.3 Transcriptional control over secondary wall fiber formation: ramifications for lignification and vascular integrity

One of the most exciting discoveries in recent years, as regards cell wall formation, is that of the roles of two transcription factors [SND1 (also called NST3) and NST1]. These are responsible for (secondary wall formation) in fibers of *Arabidopsis* (198, 199), and are specifically expressed in interfascicular fibers and xylary fibers as shown using the GUS reporter gene strategy (198, 199). Analysis of the SND1 (NST3) knockout line, however, did not initially indicate any anatomical differences when compared to the wild type, this being attributed to possible genetic redundancy. To overcome this problem, a dominant repression strategy (198) was next used where SND1 (NST3) was fused to the EAR repression domain (283). Fifteen of the 64 transgenic plants obtained displayed a phenotype that was unable to stand upright. Cross-sections of the stems showed that the interfascicular fibers and xylary fibers were very thin, due to a lack of secondary wall development; the cell walls of the vessels by comparison were unaffected.

Analysis of SND1 (NST3) and NST1 promoter activity also indicated that both transcription features were expressed in fibers suggesting that both may be involved in secondary wall thickening (199). A double mutant, *nst1-1 nst3-1*, was next generated, with the resulting phenotypes such that the plants obtained were unable to remain upright after reaching ~15 cm in length (Figure 7.13F). Stem cross-sections also indicated that the interfascicular fibers were not autofluorescent (an indication of lignification) compared to wild type, and ultrastructural analyses established the absence of secondary cell walls in the fibers; again vessels were unaffected as regards normal secondary cell wall development. Quantitative real time PCR analyses also indicated that various genes involved in secondary wall biosynthesis were suppressed (198, 199), these included those involved in monolignol (*CCOMT*, *At4CL1*, *AtOMT1*, *AtCCR1*), as well as cellulose (*irx3*, *irx5*) and xylan (*fra8*) biosynthesis. This is a very important discovery, not only that fiber secondary wall formation is under control of transcriptional factors, but also that the plant lines are again unable to compensate in any effective manner for the defects introduced.

7.7 Native lignin macromolecular configuration

It is not known with any degree of certainty as to when the biochemical pathway fully evolved to afford either the monolignols or the monolignol-derived products, the polymeric lignins, and the (oligomeric) lignans. It is presumed to have occurred with the emergence of tracheary elements, with the latter being thought to trace their origins back to about 430 million years ago (1–3). Such a lengthy evolutionary period might seem unlikely, however, to ultimately result in a random assembly process for lignins, particularly as they represent Nature's second most abundant biopolymer(s).

Yet today, there is currently still a vigorous debate regarding native lignin macromolecular configuration, namely, as to whether it results from proteinaceous control over primary structure with (presumed) template replication (29, 31, 50–52, 284–286), or whether the “random coupling” model of Freudenberg (11, 12, 53, 54) – speculated to afford lignins with 10^{66} isomers per 100-mer unit (196) – can be extended to that of “combinatorial chemistry” (175). Since such widely divergent viewpoints are apparently incompatible, determining how

native lignin macromolecular configuration is unambiguously achieved *in vivo* is essential for fully understanding the basis of vascular plant cell wall biochemistry. In this context, we considered it instructive to first briefly discuss the ten or so depictions of proposed “representative” lignin structures, together with either the evidence or lack thereof for same, and the gaps in our knowledge that remain to be filled.

7.7.1 *Early beginnings: the Freudenberg (random coupling) and the Forss (regular repeating unit) models for lignins*

Although progress was made from the 1920s to the 1960s, this was not the best period in which to scientifically study native lignin macromolecular configuration. This was not only because of the serious technological limitations encountered during this period relative to today, but also because there was no indication that phenolic radical–radical coupling could be controlled in any specific manner. Accordingly, with the best of the intentions, the structural depictions envisaged for lignins and how they were formed were highly speculative at that time to say the very least. In any event, this ultimately led to two widely divergent views of lignin structure from the Freudenberg and Forss laboratories, respectively.

7.7.1.1 *The experimental evidence: a reassessment*

It must be emphasized that one of Freudenberg’s major contributions to the study of lignins was the demonstration that coniferyl alcohol (3) coupling *in vitro*, using a “redoxase” from mushroom, afforded the racemic products (Figure 7.17A), (±)-dehydroconiferyl alcohols (68), (±)-pinoresinols (69), and (±)-*threo/erythro* guaiacylglycerol 8–O–4′ coniferyl alcohol ethers (71) in yields of *circa* 26, 13, and 9%, respectively (287). [By contrast, dehydrogenation of sinapyl alcohol (5) mainly afforded the (±)-syringaresinols (70) (288, 289).] Freudenberg then considered a gymnosperm lignin structure which was initially based on the dimeric ratios obtained for coniferyl alcohol (3) coupling, but later changed the proportions (ratios) to a predominance of 8–O–4′ interunit linkages without further experimental verification from his own laboratory (Figure 7.3A) (12, 290).

At around the same time, Freudenberg and other researchers reported that lignins were present in other living systems, such as mosses (*Polytrichum commune*, *Sphagnum*) (11, 12, 291, 292), algae (*Cystoseira barbata*, *Fucus vesiculosus*) (293–295) and possibly in fungi (e.g., *Polyporus*) (296, 297). The “conclusions” were largely a consequence of the inadequate technologies then available and the consequently misleading indications so obtained; there is no evidence that lignins are present in such organisms (298–301). It was also reported that native lignins and synthetic dehydropolymerisates were identical (53, 54), but once again this was not the case (31, 302). Freudenberg further reported that mistletoe (*Viscum album*), an angiosperm, formed a gymnosperm (G-enriched) lignin when parasitizing gymnosperms (*Pinus silvestris/Abies alba*), whereas an angiosperm (G-S) lignin was formed when it parasitized an angiosperm [white hawthorn (*Crataegus oxyacantha*)] (11, 12). This prompted him to conclude that “these are the examples of roles that lignins play in taxonomy” (11). Such claims did not, however, survive further experimental scrutiny either: Mistletoe has since been demonstrated to only biosynthesize an angiosperm (G-S) lignin through

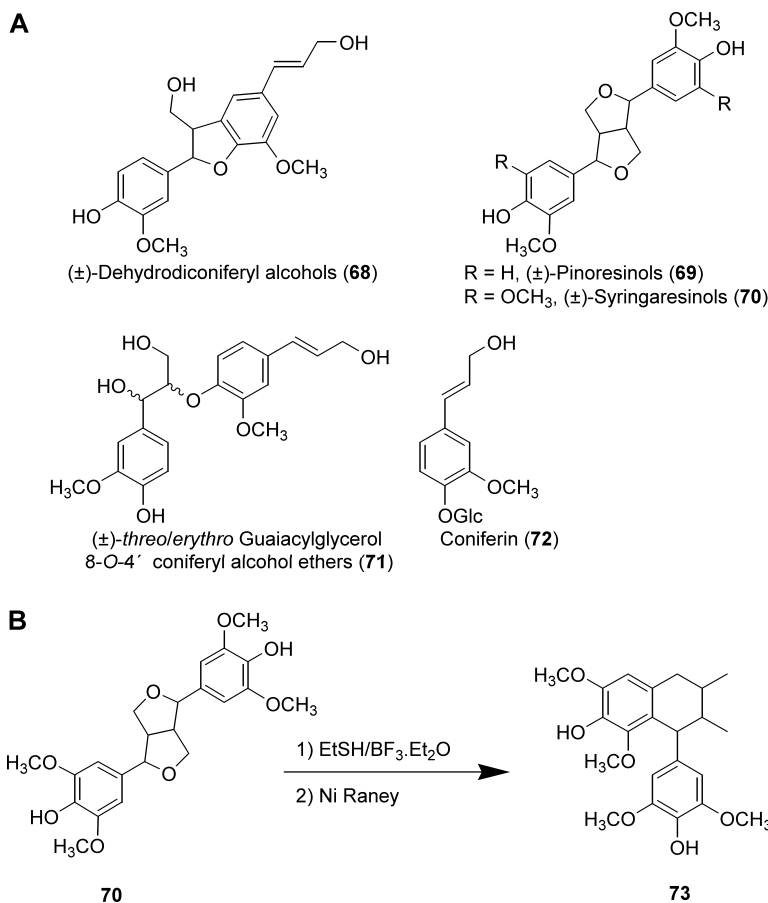


Figure 7.17 Racemic lignans **68**–**72** and (**B**) syringaresinol (**70**) rearrangement by thioacidolysis and Raney Ni treatment.

its own biosynthetic processes (303) regardless of plant source for parasitism. The findings in these studies thus serve to illustrate the quality of the data largely being obtained in the 1950s/1960s, and which accordingly contributed to the unproven notion of random assembly.

The Freudenberg laboratory also reported that monolignol glycosides (and dimeric lignans) were first formed in the cambial regions of various tree species (53, 54), and that these diffused into adjacent cells. These were then considered to be hydrolyzed back to glucose and monolignols (from the monolignol glucosides) by action of a β -glucosidase, with both the monolignols and the lignans subsequently employed for lignification (53, 54, 290). Goldschmid and Hergert (304) were unable to confirm such observations using western hemlock (*Tsuga heterophylla*), however, and later Grisebach's group (305) established that, based on turnover experiments and pool size determinations, only a part of lignin synthesis could be attributed to coniferin (**72**) metabolism. Thus, in contrast to Freudenberg's earlier assumptions, it had not in fact been determined as to how either lignin precursor transport

occurred prior to lignification, or in what form, or in what specific cells. These data, when taken together, contributed to the now long-held view of lignins being randomly assembled.

On a very different tact, experimental work undertaken on delignification of the gymnosperm, spruce (*Picea abies*) using sulphite-based chemical reagents was also carried out in the Forss laboratory. These researchers noted, based on distinct molecular size chromatographic elution profiles of lignin derivatives proper, that there were two broad classes of metabolic products solubilized (306). Forss and Fremer considered that they were derived from “hemi-lignins” and polydisperse higher molecular weight lignins, respectively, i.e., in a somewhat analogous (polyphenolic) structural relationship as for the polymeric hemicelluloses and cellulose in (woody) plant tissue. Subsequent chemical analyses of the “hemilignin” fraction by ourselves (147–150), however, established that they were instead hydrolytically cleavable mono-, di-, and tri-sulphonic acids, such as **38–41** (Figure 7.10A), etc., and were thus presumably cleaved from the native lignin macromolecule(s). That is, no evidence was obtained that they were distinct “hemilignin” biopolymers, as suggested by Forss and Fremer.

The higher molecular weight polydisperse sulphonated lignins examined by Forss and Fremer displayed some interesting differences in their chromatographic behavior, as well as in their elemental analyses, suggesting the possibility of discrete molecular species. A structural basis for these differences was, however, not actually established. Nevertheless, a speculative structure emerged from their studies in the proposal of a lignin repeating unit derived from 18 monomeric [*p*-coumaryl (**1**) and coniferyl (**3**) alcohol] entities (structure not shown), albeit without any supporting spectroscopic evidence (307, 308). Taken together, the studies by Freudenberg and Forss had thus led to two widely divergent and highly speculative depictions or views of lignins, each of which is discussed further below.

Given the limitations of the technologies and considerations about lignin formation at that time, it is perhaps not surprising that the Freudenberg depiction of lignin structure had several fatal flaws: (i) Of the 24 aromatic monomeric subunit linkages in his model (Figure 7.3A) (11), six – then revised to five in 1968 (12) – had potentially thioacidolysis cleavable monomeric units, of which one was sinapyl alcohol (**5**)-derived; gymnosperms are generally not considered to biosynthesize sinapyl alcohol (**5**) (substructure 16 in Figure 7.3A); (ii) many of the interunit linkages now known to be present in native lignins were absent in the proposed structural depiction. Absent, for example, are the potentially cleavable thioacidolysis 3–3', 3–5', or 5–5', linked dimeric entities (substructure **VII**, Figure 7.2D), as well as dibenzodioxocin (5–5', 8–O–4', 7–O–4') moieties (substructure **V**, Figure 7.2D); (iii) Eight of the 24 aromatic moieties, linking 2/3, 13a/14a, 13c/14c, and 15/17 (Figure 7.3A) also contain substructures that have never been reported as occurring in lignins proper; yet, these accounted for one-third of the linkages shown; and (iv) Between 1965 and 1968, an 8–O–4' interunit linkage was also replaced by an 8–1' substructure (11, 12).

In an analogous manner, the Forss and Fremer (307) gymnosperm lignin structure also contained similar serious structural limitations; the proposed repeating unit had eight potentially cleavable linkages [by thioacidolysis] which would also lead to monomer release, as well as a non-cleavable ten-monomeric entity containing 3–3', 8–3', 8–5', and 5–5' carbon-carbon linkages. This linkage distribution does not agree with subsequent experimental evidence obtained to the present date.

In hindsight, the depictions of lignin structures being derived from either random or non-random assembly thus lacked a sound experimental footing, and both were premature

and incorrect representations. Nevertheless, they served as a useful historical starting point, but can be eliminated from further consideration as a reasonable facsimile for either lignin structure or as an explanation as to how lignins were being formed.

At the same time, it should be emphasized that the proposed regularity in gymnosperm lignin structure was arbitrarily dismissed, with no definitive experiments, for example, being conducted and/or designed to determine the molecular basis for the intriguing molecular size chromatographic profile differences observed. More to the point, in both the Freudenberg and Forss/Fremer lignin structural depictions, experiments had not been designed to either probe lignin structure proper, or to distinguish between “random” versus “non-random” coupling, and/or to establish how lignins were indeed formed *in vivo*.

7.7.2 Further refinement of structural depictions of lignins (1970s to the present date): a reassessment

Given the lack, even today, of available technologies/methodologies in adequately probing lignin structure, there were nevertheless a number of other attempts to provide depictions of what lignin structure(s) might look like. As described below, while each contains numerous discrepancies relative to recent experimental observations, progress was made though as regards identification of additional structural subunits within lignins. However, the limitations again further underscored the urgent need to now develop methodologies to accurately probe native lignin macromolecular configuration.

7.7.2.1 1970s and 1980s: the Glasser and Glasser, Nimz, Adler, and Sakakibara models

The 1970s and 1980s witnessed several additional attempts to adequately depict possible lignin structures. The most complex structure resulted from studies by Glasser and Glasser (13), using a computerized model called SIMREL. In hindsight, this model suffered from various critical limitations: For example, it was unable to predict the presence of various substructures, such as dibenzodioxicin (5–5', 8–O–4', 7–O–4', substructure V in Figure 7.2D) moieties and it also contained a large number of hypothesized interunit linkages that have never been observed in lignins, e.g., C7–C8, C8–C7, C7–C6, and C7–O–C9. As for the Freudenberg and Forss/Fremer lignin models, hypothetical thioacidolytic cleavage of potential monomer/dimer releasable moieties in this representation would not provide quantitative data in agreement with later experimental observations, e.g., by thioacidolysis. This computer-generated model can, therefore, be eliminated from further consideration as adequately representing lignin structure.

At more or less the same time, a proposal for a representative lignin structure in the angiosperm, beech (*Fagus silvatica*), was made by Nimz (309). This model was based upon the analysis, identification, and quantification, of various monomeric, dimeric, trimeric, and tetrameric fragments released when beech woodmeal was treated with thioacetic acid and catalytic amounts of boron trifluoride at 20°C for 1 week, followed by hydrolysis of the thioacetates with NaOH (2 N, 60°C, 24 h) and Raney nickel treatment (8 h) (310). The proposed beech wood lignin structure (not shown) contained 27 inter-linked monomeric units of which, on a per monomer basis, there were 7 potentially cleavable G/S monomers,

as well as 8–1' (4), 8–5' (2), 8–8' (1.5), 7–8' (1), 5–5' (1), and 4–O–5' (0.4)/7–6' (0.1) dimers. In hindsight, some of these fragments may have resulted from rearrangement, e.g., of syringaresinol (substructure **IIIC** in Figure 7.2D) during the chemical degradative procedures employed, as observed for thioacidolysis cleavage/Raney Ni treatment where syringaresinol (**70**) can be converted into the tetrahydronaphthalene derivative **73** (Figure 7.17B) (217, 311) (Jourdes *et al.*, unpublished results). Furthermore, the linkage frequencies in the Nimz model are not fully consistent with data obtained for other angiosperm lignins, such as with *Arabidopsis* and alfalfa, this perhaps being due to low chromatographic recoveries of various products, etc. Other linkages in the Nimz model, such as 7–8', are also not known to be present, and may thus potentially represent artifacts as well. As before for the other lignin structures discussed, various other known subunit structures, such as 5–5', 8–O–4', 7–O–4'-dibenzodioxocin, were absent from this model.

Two other “representative” structures for gymnosperm lignins were reported by Adler (312) and by Sakakibara (313). The first (not shown) contained a proposed spruce lignin depiction of 16 aromatic monomer (H/G) units, with 5 potential thioacidolysis releasable monomers, including one again being the unlikely sinapyl alcohol (**5**). Other linkages included thioacidolysis cleavable 8–5' (1), 5–5' (1), and 8–6' (1) dimeric entities, with the remaining five monomers envisaged to be linked together via 8–1', 4–O–5', 8–8', and 5–5' interunit linkages. Sakakibara (313), by contrast, proposed another “representative” softwood (gymnosperm) lignin structure (also not shown), based upon hydrolysis and hydrogenolysis analyses. This contained 35 monomeric aromatic units which could be potentially cleaved (e.g., by thioacidolysis) to afford G/S monomers (7), 8–1' dimers (3), 8–5' dimers (3), 8–8' dimers (1), trimers (3) consisting of 5–5'/4–O–5', 8–8'/5–5', and 4–O–1'/8–5' linkages, as well as a proposed five-unit (8–5', 8–5', 5–5', 8–8') linked moiety. None of these proposed structures, however, again apparently adequately account for gymnosperm lignins, in terms of interunit type and frequency. Nor did they contain the additional substructures more recently discovered, such as the 5–5', 8–O–4', 7–O–4'-dibenzodioxocin (**V**, Figure 7.2D). Thus, while recognizing all of these to be valuable studies, they fell short of being adequate representations of native lignin structure(s).

7.7.2.2 1990s and 2000s: the Brunow and Banoub/Delmas depictions of lignin structure

With the incremental advances made in the study of lignins, such as by application of NMR and mass spectroscopic analyses, together with thioacidolytic degradation, further refinement of possible lignin representations were attempted for these polymers in spruce (15) and wheat straw (314). The Brunow *et al.* (15) model speculated the existence of random-linked structures, such as the 28-unit structure shown (Figure 7.3C), and also that dibenzodioxocin (substructure **V**, Figure 7.2D) and 5–5' linkages (substructure **VII**) served as important branching points. The envisaged structure (Figure 7.3C), while acknowledging herein its hypothetical basis as emphasized by the researchers themselves, would nevertheless contain on a per monomer basis 10 potentially cleavable thioacidolysis monomers (35%), with the remaining eighteen units (65%) being releasable as 8–5' (3), 5–5' (2), 8–1' (2), 4–O–5' (1), and 8–8' (1) linked moieties – where one of the latter has a presumed C-5 linkage to an adjacent lignin subunit, respectively. As gleaned from inspection of this

proposed “hypothetical” structure, much of it would presumably be readily susceptible to thioacidolysis degradation. However, release of such fragments/substructures in the relative amounts experimentally determined has not been observed. Furthermore, nor has it been established that either the dibenzodioxocin or 5,5′ substructures serve as branching points. Thus, this proposed structure again does not meet experimental scrutiny.

A linear (8–5′) linked lignin macromolecule has also been proposed (314), based on mass spectrometric analysis (APCI-MS, MS/MS, and MALDI-TOFMS) of the extracted lignin from wheat straw using the AVIDEL (315) procedure. Such proposed structures, however, need to be verified, quantified and placed in context with the existing chemistry of lignins, in order to assess what, if any, their relative merits and contributions are to lignin macromolecular configuration. Whatever the limitations of this approach, these researchers sought to obtain needed primary sequence data, unlike many of the previous highly speculative lignin structural models which did not.

7.7.3 A new beginning: the need to fully define native lignin macromolecular configuration proper

As indicated above, none of the various “representations” of lignin structure have adequately reflected native macromolecular configuration. Indeed, there may not be any other natural product whose “structure(s)” have been approximated through attempts to determine subunit linkage type and frequency, but not through obtaining sequences of the interunit linkages. However, the trends noted through attempts to obtain precise, rigorous, quantification and identification of subunit type, and their frequencies, in lignins present in *Arabidopsis* and alfalfa are considered indicative, at least by ourselves, of non-random assembly. This is not to say though that all interunit linkages can yet be accounted for – currently, ~40–45% can at best be fully quantified using these methods. In addition, lignin analyses suffer from another limitation: Technologies have not yet been developed to probe precise lignin structures in different cell wall layers and/or in distinct cell wall types (e.g., xylem versus fibers).

7.7.3.1 Limited substrate degeneracy during dehydrogenative polymerization in vivo versus random/combinatorial chemistry

The systematic analyses of various transgenic/mutant lines has identified important trends in lignin macromolecular assembly. These analyses (71, 72, 131, 132, 215) have also eliminated non-monolignol entities, such as feruloyl tyramine (**60**), acetosyringone (**61**), etc. as being involved in core lignification in contrast to previous assertions (173, 174). Additionally, as summarized in Figure 7.14, modulation of the monolignol-forming pathways in the same organisms gave H, G, 5OH-G, and S-enriched lignins whose subset of identifiable, cleavable, interunit linkages were apparently invariant of hydroxylation/methoxylation patterns of aromatic ring substitution. In the case of H-monolignol deposition, however, this was prematurely terminated at a “metabolic checkpoint” (72, 215), the reasons for which need to be established. A similar situation also held for the *poly-p*-hydroxycinnamaldehydes (71),

with severe adverse effects being noted on plant structure overall and thus on vascular integrity. For the H, G, 5OH-G, and S enriched lignins, this subunit invariance would not be expected a priori as the H, G, 5OH-G, and S monolignols differ in having from 5 (H) to 3 (S) potential sites available for radical–radical coupling. Additionally, the catechol nature of 5-hydroxyconiferyl alcohol (**4**) represents yet another potentially confounding feature, as discussed earlier.

Yet the data obtained in lignin subunit characterization and frequency suggest a limited substrate degeneracy during the dehydropolymerization step, i.e., whereby the amounts of (cleavable) 8–O–4' and 8–1' interunit linkages present are kept directly proportional to lignin content, but apparently invariant of hydroxyl/methoxyl group aromatic ring substitution pattern. Moreover, for the H- and G-enriched lignins, the quantifiable amounts of releasable 8–5', 5–5', and 5–O–4'-derived dimers followed similar trends, whereas neither the 8–8' linked ligballinol (**58**) and pinoresinol (**69**) were released in any appreciable amount. The latter is likely consistent with their presumed covalent modification during native lignin formation, perhaps at C-5. Analyses of the S-enriched *Arabidopsis* line also provided useful insights: While only ~32% of the linkages could be accounted thus far via thioacidolysis, the 8–8' linked syringaresinol (**70**) subunits were only present in very small amount as their thioacidolysis derivatives. These data suggest that the S-enriched lignin was not composed primarily of S-units linked through 8–O–4' and 8–8' interunits as previously envisaged (i.e., substructures **Ic** and **IIIc**, Figure 7.2D), the reasons for which need to be determined. Additionally, the COMT knockout line (in *Arabidopsis*) resulted in an apparently equivalent reduction of G/S monomeric moieties leading to benzodioxane formation. Such data, for the COMT knockout mutant line in *Arabidopsis*, provisionally suggest that in the cell wall types normally forming S-units, each of the 5-hydroxyconiferyl alcohol (**4**) moieties is linked to a coniferyl alcohol (**3**) moiety; this is again envisaged to place another severe limitation on how native lignin macromolecular configuration is achieved.

Taken together, the data have nevertheless begun to provide useful insight into the trends involved in lignin macromolecular assembly, and give an impetus to explore and determine lignin primary structure. We do not consider, however, that these data are consistent with a “random assembly/combinatorial chemistry” model with 10^{66} or more isomers, but instead reflect limited substrate degeneracy during native lignin formation, as is observed in many areas of metabolism. It is presumably significant that several of the preceding enzymes in the monolignol-forming pathway (4CL, CCR, and CAD) are also substrate versatile. This, in turn, may have an important bearing on the limited substrate degeneracy observed for lignin macromolecular assembly.

7.7.3.2 Lignin properties: evidence for polymer cross-links and branch points, lignin association and template polymerization?

Beginning with the studies of Freudenberg and others, it was initially assumed that lignins were three-dimensional cross-linked biopolymers, albeit also in the absence of any experimental data in support of this assumption. Such assumptions, however, did not survive further experimental scrutiny either. In a study by Dolk *et al.* (316), using kraft lignin derivatives, it was estimated that at best there was less than one cross-link per 19 monomeric units. Later studies by Mlynár and Sarkanen (317) and by Smith *et al.* (318), using ultracentrifugation/

size exclusion chromatography and modeling of delignification with a computer program (SIMREL), respectively, also concluded that there were no detectable cross-links. Moreover, other studies probing lignin biophysical properties in situ have indicated that lignins were more akin to linear-like polymeric entities (319). Interestingly, Forss and Fremer (308) had underscored many of the experimental shortcomings in the proposed Freudenberg three-dimensional cross-linked structure which went unanswered. It is thus quite bewildering why many researchers have arbitrarily dismissed the concept of regular structure in lignins without the appropriate experiments having even been either designed and/or conducted. Yet this incorrect notion of three-dimensional cross-linked polymer for lignins, however, exists even today in many literature contributions.

Lignins also display a profound tendency to associate, as demonstrated, for example, in the study of kraft lignins (8). Specifically, it has been reported that kraft lignins – while chemically modified but which nevertheless are presumed to retain important vestiges of native lignin structure – contain discrete molecular entities. These can associate with one another to form multimodal distributions of interconvertible supramolecular complexes estimated to be comprised of 10^3 – 10^4 individual species. Such selectivity is considered to result from association of well-defined (regular) structures in the participating polymer chains (320). This presumably would not be expected if random (10^{66} /100-mer) and/or combinatorial chemistry was occurring.

In presumed agreement with a concept of defined template polymerization, one-electron oxidation of coniferyl alcohol (3) moieties in open solution in the presence of a lignin template apparently engendered preferential formation of high molecular weight species in preliminary studies. In this regard, Sarkanen *et al.* (50, 51) reported the effects of polymerization of coniferyl alcohol (3) in the presence of peroxidase/ H_2O_2 and a methylated “kraft” lignin preparation (Mw $\sim 15\,400$, 2.7×10^{-8} M initial concentration). In the presence of the putative template, higher molecular weight entities were reportedly preferentially formed as shown by the elution profiles (Figure 7.18A), and to a much lower extent in absence of the template (Figure 7.18B). These researchers interpreted these findings as due to template polymerization effects, where the preformed (methylated kraft) lignin macromolecule assumed the role of a progenitorial template in vitro. That is, monolignol radicals were considered to be positioned on adjacent loci of the template with new interunit linkages determined by either the corresponding substructure in the lignin template chain and/or the chemical nature of the monolignol (radical) species aligned on the template. Work has, however, not yet been described as to the chemical (subunit) nature of these interunit linkages and how/if the fidelity of the replication process is maintained relative to that of the macromolecular template itself. Nevertheless, the presumed ability to polymerize monolignol radicals on a preexisting lignin template needs to be considered as regards possible relevance to cell wall assembly mechanisms.

7.7.3.3 *Monolignol radical-binding proteins: dirigent proteins in phenolic coupling – monolignol to terpenoid metabolism?*

We initially considered it instructive to determine how control over monolignol radical coupling might be effectuated, particularly given that this was a factor largely not considered as being possible during the 1950s/1960s. On the other hand, this was a relevant issue with the

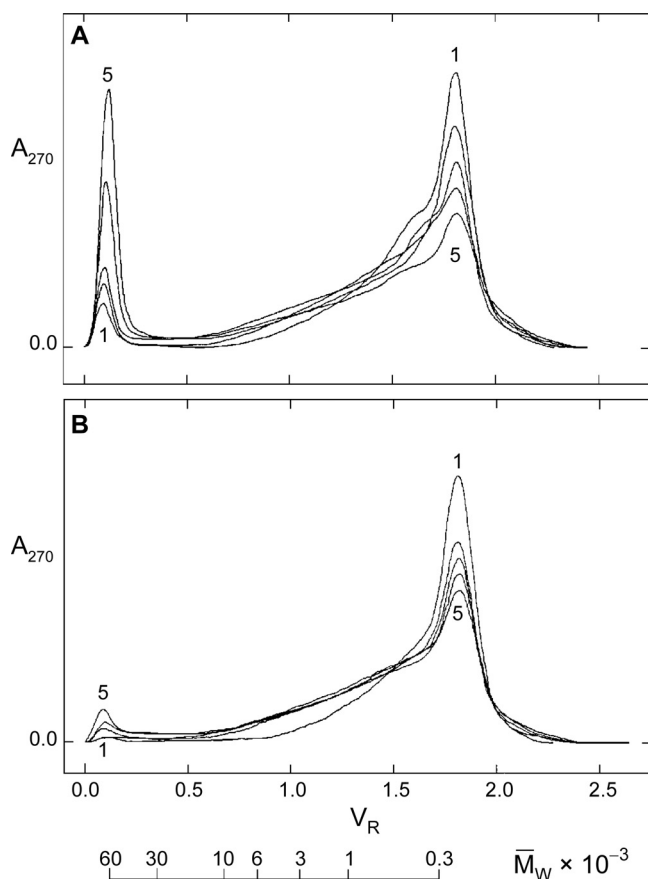


Figure 7.18 Molecular weight distributions of dehydropolymerisates successively formed under limiting Zutropfverfahren conditions from monolignol, coniferyl alcohol (3) in (A) presence and (B) absence of methylated macromolecular lignin template after (1) 20 hours, (2) 50 hours, (3) 70 hours, (4) 75 hours, and (5) 80 hours. (Sephadex G100/aqueous 0.10 M NaOH) (50). (Reprinted from *Phytochemistry*, vol. 45, Guan, S.Y., Mlynár, J. & Sarkanen, S., Dehydrogenative polymerization of coniferyl alcohol on macromolecular lignin templates, pp. 911–918, Copyright 1997, with permission from Elsevier.)

large numbers of structurally related lignans now known. In this regard, many different forms of specific coupling have been reported in isolated metabolites (e.g., containing specifically-linked 8–1', 8–5', 8–O–4', 5–5', 3–O–4', 7–1', 8–7', 1–5', and 2–O–3' interunit linkages) depending upon the metabolite and/or plant species in question (321, 322). This indicated, at the very least, that a mechanism of regiospecific coupling control had evolved *in planta* for phenolic coupling. Additionally, large numbers of lignan metabolites are optically active suggesting, in turn, that stereoselective coupling might also be occurring in many instances. For the purposes of this discussion, however, there are two pertinent examples of potential control over phenoxy radical–radical coupling. These include formation of the 8–8'-linked (+)-pinoresinol (69a, Figure 7.19A) in *Forsythia* species (323–327), and that of (*meso*)

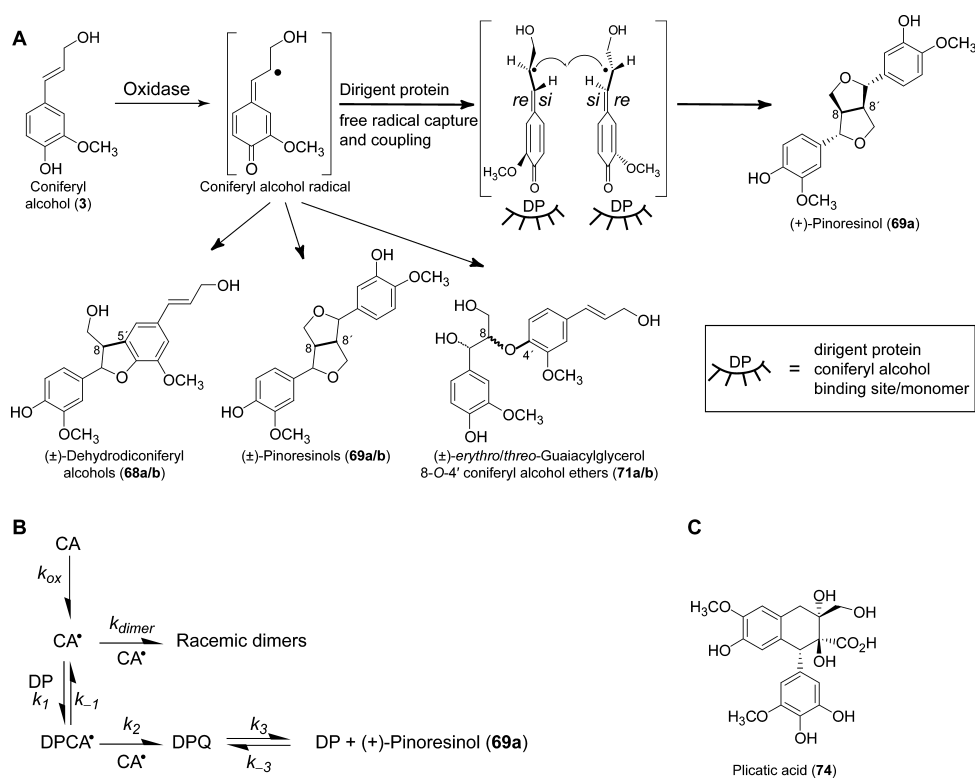


Figure 7.19 (A) Formation of dirigent protein (DP)-mediated stereoselective coupling versus that of non-specific (racemic) coupling of coniferyl alcohol (3). (B) Proposed kinetic model for dirigent protein (327). CA, coniferyl alcohol (3); CA*, coniferyl alcohol (3) radical; DP, dirigent protein; DPCA*, dirigent protein–coniferyl alcohol (3) radical complex; DPQ, dirigent protein quinone methide intermediate complex; k_{ox} , rate constant of coniferyl alcohol (3) oxidation; k_1 forward rate constant of coniferyl alcohol (3) radical binding to DP; k_2 , rate constant of coniferyl alcohol (3) radical binding to DPR complex; and k_3 , rate constant of (+)-pinoresinol (69a) release. (C) Plicatic acid (74).

8–8'-linked nor dihydroguaiaretic acid (84, Figure 7.20C) in the creosote bush (*Larrea tridentata*) (328, 329), respectively.

Work was thus undertaken to initially establish how the lignan (+)-pinoresinol (69a) was formed, this having been an earlier unresolved scientific interest of Holgar Erdtman. Using *Forsythia suspensa* as a biological partner, we were able to demonstrate that preferential stereoselective coupling of two *E*-coniferyl alcohol (3) moieties occurred to afford (+)-pinoresinol (69a) when incubated with crude “cell wall”/insoluble preparations (323). Subsequent purification of the various proteinaceous components in this crude preparation ultimately afforded (as estimated by SDS-PAGE) an ~110 kDa one-electron oxidase (a laccase) and an ~26 kDa protein, with the latter protein lacking monolignol oxidizing capacity (324). The laccase alone in vitro generated coniferyl alcohol radicals which underwent the well-known nonspecific coupling to afford the corresponding racemic (±)-dihydrodiconiferyl alcohols (68a/b), (±)-pinoresinols (69a/b), and (±)-threo/erythro guaiacylglycerol 8–O-4'

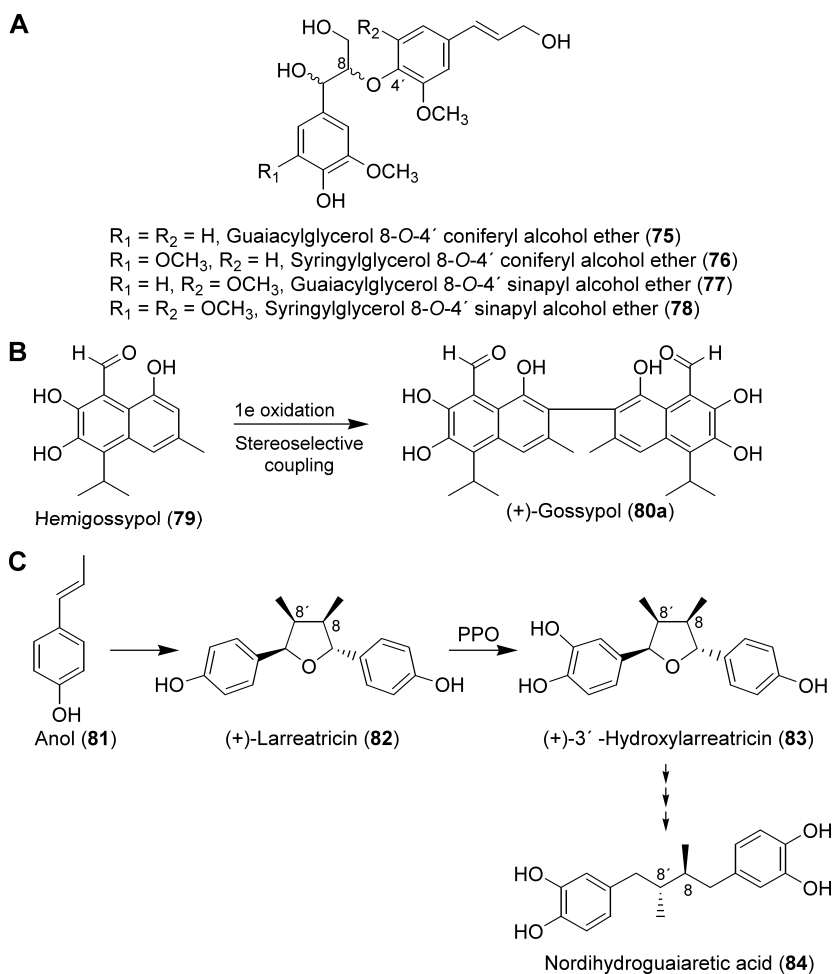


Figure 7.20 Biochemical conversions proposed thus far for: **(A)** diastereoselective 8-*O*-4' homo-/hetero-coupling of coniferyl (**3**) and sinapyl (**5**) alcohols in *Eucommia ulmoides* to give lignans **75–78** (335, 336); **(B)** stereoselective coupling of the achiral hemigossypol (**79**) in the presence of a presumed dirigent protein from (*Gossypium hirsutum* L. var. *marie galante*) flower petals and a laccase to afford (+)-gossypol (**80a**, stereochemistry not shown) (337), and **(C)** nordihydroguaiaretic acid (**84**) formation from the presumed achiral precursor, *p*-anol (**81**).

coniferyl alcohol ethers (**71a/b**) (Figure 7.19A). On the other hand, when the 26 kDa protein, which existed as an ~50–52 kDa dimer, was added to the assay mixture, stereoselective coupling occurred instead to afford (+)-pinoresinol (**69a**) (Figure 7.19A).

Interestingly, in addition to the laccase, other one-electron oxidases (peroxidase) and one-electron oxidizing agents were able to effectuate stereoselective coupling in the presence of this protein (324). Given this striking ability to dictate the outcome of stereoselective coupling with the (+)-pinoresinol-forming protein, we coined the term dirigent protein (DP) from the Latin: *dirigere*, to guide or to align (324). The corresponding gene was next

obtained, with this encoding a protein of ~ 18 kDa. The discrepancy in the molecular size (~ 26 versus ~ 18 kDa) was due to posttranslational (glycosylation) modification (325); functionally competent recombinant protein was also obtained when expressed in a “glycosylated” form using insect (*Spodoptera frugiperda*) cell cultures (325).

The biochemical mode of action of the (+)-pinoresinol forming DP has since been the subject of our recent work (326, 327), where it is considered to function by trapping monolignol (radicals). Figure 7.19B thus depicts our current understanding of the kinetic and “rate-limiting” processes presumed to be involved in stereoselective coupling, versus that of nonspecific coupling leading to racemic dimers (327). Importantly, it was demonstrated that proteins had indeed evolved monolignol (radical) binding capacity, and the ability to engender specific coupling modes, in vascular plants.

However, we considered the (+)-pinoresinol forming DP discovery as a special example of control over monolignol radical–radical coupling, i.e., whereby the monolignol (radical) was bound to the active site of each monomer in the DP dimer. In this way, the (+)-pinoresinol-forming DP dimer was able to orientate both coniferyl alcohol radicals in such a way as to where only (+)-pinoresinol (**69a**) formation could occur (Figure 7.19A).

More recent work has established that dirigent proteins and their homologues are found throughout the vascular plant kingdom (330–332). However, they appear to be restricted to land plants, suggesting they obtained their function(s) during the transition of plants to land. Dirigent proteins are generally also present in large multigene families with varying levels of homology (e.g., from 99.5 to 12.5% identity) (331, 332), with most biochemical/physiological functions of their homologues currently unknown. As a beginning to define their functions, we have investigated their expression profiles, using a GUS-reporter system linked to each putative DP promoter of two dirigent multigene families and/or homologues in both western red cedar (*Thuja plicata*) (333) and *Arabidopsis* (334) (Kim *et al.*, manuscript in preparation). [Western red cedar was of particular interest since it accumulates the (+)-pinoresinol-derived metabolite, plicatic acid (**74**, Figure 7.19C), and several of its DPs have the capacity for (+)-pinoresinol (**69a**) formation (330).] Determination of their expression profiles was carried out for both families; this appears to be a useful approach to begin to establish the biochemical/physiological functions of the various homologues.

There have also been a number of other studies which provisionally suggest DP control over other coupling modes. For example, using *Eucommia ulmoides* “insoluble residues” Lourith *et al.* (335, 336) reported diastereoselective homo- and hetero-coupling of sinapyl (5) and coniferyl (3) alcohol moieties in vitro, without addition of any cofactor, to afford the 8–O–4' lignans **75–78** (Figure 7.20A) of differing levels of *erythro/threo* ratios and optical activities (335, 336). The proteins involved in such coupling now need to be purified, and their encoding genes cloned, in order to more fully characterize the basis of such transformations.

Another intriguing report of presumed dirigent protein involvement is in the stereoselective coupling of two molecules of the terpenoid, hemigossypol (**79**), to afford (+)-gossypol (**80a**) in cotton (*Gossypium hirsutum* L. var. *marie galante*) flower petals (Figure 7.20B) (337). Interestingly, different *Gossypium* varieties accumulate varying levels of (+)- and (–)-gossypols (**80a** and **80b**), with both having equal toxicity toward insects and pathogens. However, cotton seeds generally cannot be used in animal feed because of the toxicity of (–)-gossypol (**80b**). Breeding has thus been used to select varieties accumulating the (+)-, but not the (–)-, antipode of gossypol (**80**). Research has also recently been conducted to

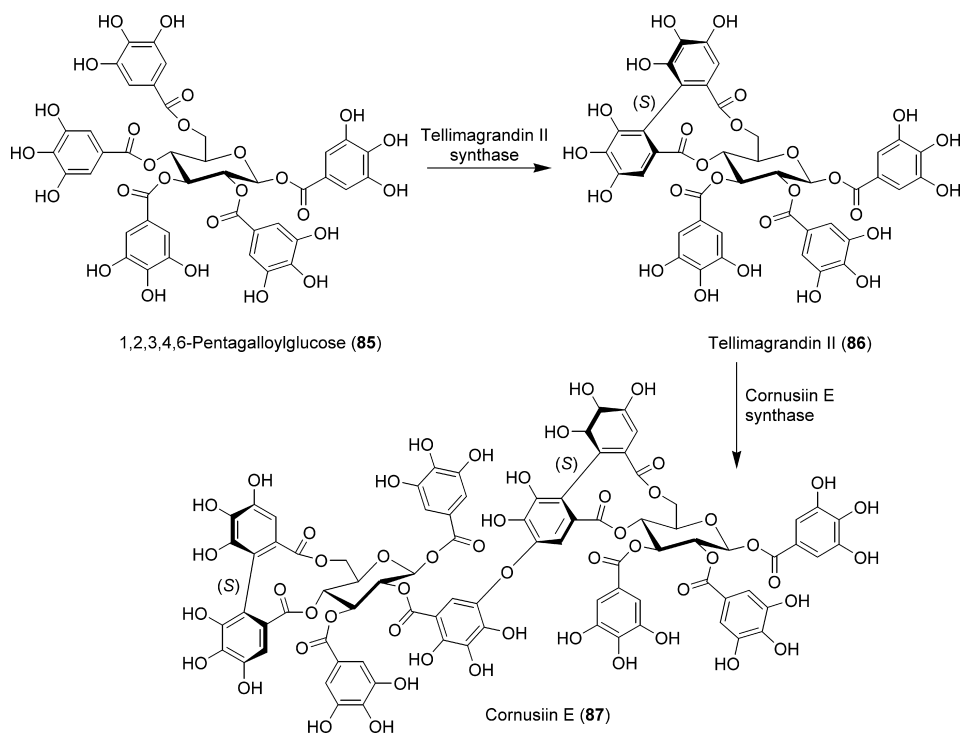


Figure 7.21 Phenolic coupling in *Tellima grandiflora*: Intermolecular coupling of 1,2,3,4,6 pentagalloyl glucose (**85**) to afford tellimagrandin II (**86**) and intra-molecular coupling of the latter to give cornusiin E (**87**) (338–340).

understand the biochemical basis of (+)-gossypol (**80a**) formation in cotton flowers (337). In the presence of a one-electron oxidase (peroxidase, laccase), coupling of hemigossypol (**79**) only afforded racemic gossypol (**80a/b**), whereas when a presumed dirigent protein – again lacking oxidative capacity was added – stereoselective coupling occurred to essentially only give (+)-gossypol (**80a**). Thus, it would provisionally appear that DP control over radical–radical coupling can involve metabolically quite distinct plant product classes.

Other forms of control over regiospecific radical–radical coupling have also been noted for formation of the 8–8'-linked lignans, such as in the creosote bush (*Larrea tridentata*). The latter accumulates (*meso*)-nordihydroguaiaretic acid (NDGA, **84**) and other 8–8'-linked lignans. Interestingly, while the presumed precursor of NDGA, *p*-anol (**81**) can potentially undergo various forms of coupling at different sites on the molecules, such as at the 4-*O*, C-5, C-1, and C-8 positions, only regiospecific coupling at 8–8' occurs (see Figure 7.20C). This is again indicative of, at the minimum, regiospecific control over phenolic radical–radical coupling (328, 329). Other examples of regiospecific coupling are also apparently found in ellagitannin metabolism, such as in the intermolecular coupling of pentagalloyl glucose (**85**) moieties to afford tellimagrandin II (**86**) and subsequent intramolecular coupling of the latter to generate cornusiin E (**87**) (Figure 7.21) (338–340). While these are presumably considered to be “laccase-like” protein mediated (338–340), it will be important to establish how their biochemical mechanisms differ – if they do – from the dirigent

protein-mediated coupling. Work is currently underway to explore this possibility. In any event, biochemical mechanisms have been preliminarily described for proteinaceous control over both stereoselective and regiospecific coupling.

7.7.3.3.1 PROTEIN VERSUS NON-PROTEIN DIRECTED NATIVE LIGNIN MACROMOLECULAR CONFIGURATION AND THE QUESTION OF RACEMISATION IN LIGNIN STRUCTURE

The preceding sections have emphasized some of the difficulties in establishing native lignin macromolecular configuration – beginning with the technological limitations experienced in lignin analyses more than five decades ago – to the present date. In addition, verification of the long-known racemic nature of lignins and of lignin subunit fragments apparently convinced some researchers more recently that lignin formation must indeed occur randomly in vivo (175, 341). However, the presence of racemic substructures does not eliminate proteinaceous control over lignin macromolecular configuration as described below. Additionally, the notion of random coupling has in turn led to other suggestions – but again without rigorous proof – that various cell wall constituents, such as hemicelluloses, cellulose, have important roles in determining and/or establishing lignin configuration in vivo. Several of the hemicelluloses may, however, indeed be involved in forming lignin–carbohydrate bonds, e.g., via reaction with intermediate quinone methides.

However, such considerations fully ignore both the presence and the roles of cell wall proteins, the vast majority of whose functions remain currently unknown (342). In this regard, *all* of the advances made to the current day in the study of phenylpropanoid metabolism (and the effects of its modulation) resulted solely from the study of the proteins, enzymes, and genes involved. Thus, given the paucity in our knowledge of cell wall biochemistry (and of the proteins involved), we consider it inopportune not to systematically examine their roles in controlling native lignin macromolecular configuration. Indeed, this is a more likely and presumably more productive direction than the study of non-proteinaceous components, such as the effects of cellulose, hemicelluloses, etc.

In this regard, various proteins have been considered for their potential roles in establishing lignin macromolecular configuration in vivo. For example, using polyclonal antibodies raised against the (+)-pinoresinol-forming dirigent protein, our preliminary analyses indicated that their epitopes could be detected in the cell wall areas (e.g., S1 sublayers and cell corners) where lignification was initiated (284). This was considered as being due to recognition of the monolignol (radical) binding motif(s). This, in turn, led to our proposal that – in contrast to stereoselective coupling – there were arrays of dirigent (monolignol) radical binding sites in those subcellular regions, thereby providing the basis for forming a predetermined – albeit racemic – lignin structure (or structures) (29, 31, 284, 285).

Other studies have also implicated various other proteins (e.g., proline-rich proteins, PRP) as potential “lignin scaffolds” in the cell wall, based on co-localization of PRP epitopes and lignins in developing cell walls of maize coleoptiles (343) and in secondary cell walls of soybean (*Glycine max*) differentiating protoxylem elements (344), albeit without any precise indication as to what this meant either biochemically or in terms of how they influence lignin macromolecular configuration. Before investigating the involvement of any one of these possibilities, we considered it useful to begin to develop more robust approaches to: probe native lignin structure in vivo (type and frequency of interunit linkages); identify conditions for obtaining native lignin facsimiles through in vitro assays, as well as to identify

the biochemical (structural motifs) in dirigent proteins that are required for monolignol (radical) substrate binding (work in progress).

Proponents of the random coupling/combinatorial biochemistry model – affording 10^{66} (per 100-mer) isomers (196) alone for monolignols 1-, 3-, and 5-derived structures – have given additional speculation as to why it is their consideration that no proteinaceous control over lignification is in effect. These include: (i) the presumed absence of optical activity in lignins; (ii) that for proteinaceous control, there would be a requirement for complementary chains of proteins in both D and L configurations (of their amino acids) for monolignol (radical) binding leading to racemic lignins; and (iii) that presumed lignin subunits, such as the ~1% or so of secoisolariciresinol components considered part of gymnosperm lignin, are formed nonenzymatically.

In this regard, in 1999, Ralph *et al.* (341) endeavored to demonstrate that the presence of racemic lignin fragments eliminated the possibility of proteinaceous control over macromolecular lignin configuration. This was an unexpected interpretation, given that provisional reasoning had been provided beforehand (31) and later (52) to rationalize the hitherto well-known absence of lignin optical activity. Nevertheless, presumed 8–5', 8–1', 8–8', and 8–O–4' lignin fragments were isolated from pine sapwood, using the reductive DFRC method (341), which converts benzylic hydroxyl groups of various lignin/lignan-derived entities to their corresponding methylenic functionalities. Analyses of these products established that two (8–5' and 8–1'-derived) reduction products were, as expected, racemic; on the other hand, the reduced derivatives obtained from the presumed 8–8' pinoresinol (**69**) and the 8–O–4' (**71**) substructures could not be resolved into specific enantiomeric forms. Interestingly, the 8–8'-linked product was not pinoresinol (**69**) *per se*, but was instead secoisolariciresinol (**89**) (tetraacetate): as indicated above, the latter **89** is currently considered by some researchers as being a minor (~1%) component of gymnosperm (spruce) lignin (345). [Indeed, the abundance of this presumed structure was estimated to be ~1.0 unit per 100 C9 units in spruce lignin, and 1.0–1.5 units per 100 C9 units in kraft and kraft pulp residual lignin based on quantitative ^{13}C NMR and HSQC NMR analyses (345).]

Formation of secoisolariciresinol (**89**) from pinoresinol (**69**) (Figure 7.22A), however, has been extensively investigated in this laboratory, this resulting from the action of either enantiospecific pinoresinol–lariciresinol reductases (PLRs) or with PLRs that do not display strict enantiomeric preferences (346–348). Moreover, the X-ray crystal structure of PLR has been obtained (348), a catalytic mechanism proposed (348), and work has been completed to identify the changes in PLR substrate-binding pockets for differing enantiospecificities/preferences (Kim *et al.*, manuscript in preparation). It is now quite well established that enantiospecific differences and/or slight enantiomeric preferences involve, as anticipated, changes in the nature of the substrate-binding pocket(s). That is, this does not require an interconversion of the D- and L-configurations of the amino acids present in the protein/enzyme structure, as proposed by Ralph and Brunow (349). In a somewhat analogous manner, the 7–O–4' reduction of (\pm)-dehydrodiconiferyl alcohols (**68a/b**) by phenylcoumaran benzylic ether reductase (PCBER) results from binding either enantiomer in its large substrate-binding pocket (348), i.e., again without any necessity to have D and L forms of the amino acids in the enzymes involved.

Interestingly, Holmgren *et al.* (350) also attempted to demonstrate the reductive conversion of pinoresinol (**69**) into secoisolariciresinol (**89**) upon incubation of coniferyl alcohol (**3**) with horseradish peroxidase/ H_2O_2 in presence of NADH, i.e., in the absence of any

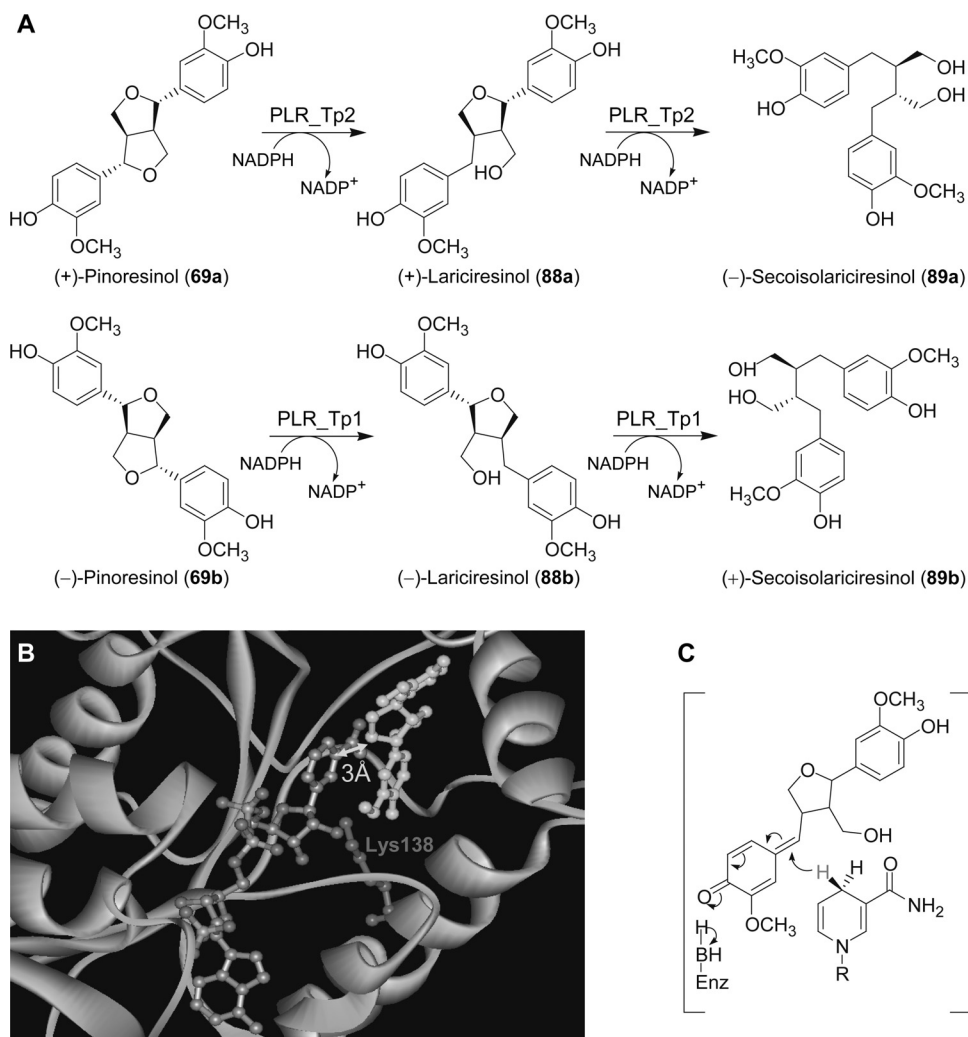


Figure 7.22 Pinoresinol-lariciresinol reductases (PLRs)/PLR homologues, PLR_Tp1 and PLR_Tp2, in the gymnosperm western red cedar: **(A)** Differing enantiospecificity differences of distinct PLR (homologue); **(B)** Partial crystal structure of PLR_Tp1 showing general catalytic base, Lys138, together with substrate (**69b**) and NADPH; and **(C)** Proposed sequential reduction to secoisolariciresinol (**89**) via intermediary quinone methide. (Reproduced in color as Plate 23.)

(PLR) protein. As expected, no such reduction occurred in the absence of functional PLR; indeed, such experiments are generally carried out as controls in our enzyme assays. Thus, given that biochemical mechanisms for the reduction of both (+)- and (-)-pinoresinols (**69a** and **69b**) have been described (347), it may be more instructive to examine whether such processes are involved – even to a small amount – in either the lignification process, or whether they result from an infusion of heartwood-forming components, including small amounts of (±)-secoisolariciresinols (**89**).

7.8 Future outlook: remaining questions in lignin macromolecular assembly/configuration, proposed lignin template replication, and overall cell wall formation

The fields of lignin chemistry, biochemistry, molecular biology, and plant cell wall biomechanics/anatomy are of greatly renewed interest as the twenty-first century unfolds. This contrasts with the preliminary studies of the mid-twentieth century on lignin constitution, which more or less relegated this important class of macromolecules to the scientific curiosity of being randomly assembled. This, in turn, was largely as a result of the severe technological barriers that existed during the 1950s and 1960s, and the speculations of the time. Yet, even today, there remains spirited discussion as regards both lignin structure and how they are formed.

From the standpoint of the authors, the recent advances made largely in the last decade or so now beckon a different viewpoint of the lignins – particularly as regards their potential to help facilitate either cellulose-to-ethanol or related biofuels/bioproducts technologies. This is thus a very exciting time for studying lignins, as there is still much to be learned about both lignins and the associated cell wall assembly processes. In addition, such problems need to be solved to meet humanity's needs for future generations.

Several areas that seem to be of particular promise for future study are summarized below:

- 1 *Monolignol Transport to the Cell Wall*: Establishing the biochemical basis of how monolignols **1**, **3**, and **5** (or derivatives/homologues thereof) are targeted to precise regions within the cell walls, and how this is regulated, are important questions to resolve; while (ABC) transporters (351) have been postulated as involved, this has not yet been proven.
- 2 *Lignin Initiation Sites*: Lignin deposition begins at so-called “initiation sites” in the cell corners and then progresses to eventually encapsulate the entire cell wall prior to apoptosis (see Figure 7.5). In this regard, establishing the nature and function of each of the proteins, enzymes and genes involved at these initiation sites appears to be a particularly attractive line of enquiry. Resolution to this should substantially clarify any remaining questions on how lignins are formed.
- 3 *Lignin Primary Sequences/New Chemistries*: A related need is to develop methodologies to establish the primary sequences of the lignins being formed initially, as well as developing new technological means and new chemistries to both identify and quantify all interunit linkages. This will also include establishing unambiguously whether lignins contain either no branches, or short or long branches, and how lignin–carbohydrate linkages are both formed and regulated.
- 4 *Re-oxidation of the Growing Lignin Chains*: Another urgently needed emphasis is in establishing biochemically how monolignols **1**, **3**, and **5** and the growing lignin chains are re-oxidized in the cell wall following initial coupling, i.e., thereby enabling radical/radical generation to reoccur. This is a problem that has long been recognized, namely, how does one-electron re-oxidation of lignin continue at sites presumably distant from the one-electron oxidative enzymes. Whether this occurs, for example, via some form of electron transfer through the lignin matrix (e.g., originating via oxidation by the presumed peroxidase/H₂O₂) or whether a diffusible oxidant, such as has been suggested for Mn³⁺ is involved (352), needs to be established.

- 5 *Monolignol Radical and Lignin Primary Chain Interactions, Proposed Template Polymerization, and Lignin Association*: Unambiguously determining whether the monolignol radicals being generated are never “free,” but instead are transiently immobilized as a result of strong $\pi \dots \pi$ interactions between the substrates and a pre-existing lignin macromolecular template is also another important goal, i.e., following on the previously proposed template polymerization for replication of lignin primary chains (353). Recently, Sarkanen and Chen modeled non-covalent interactions between a monolignol (coniferyl alcohol, 3) radical and a representative monomer residue (= veratryl alcohol) for a lignin macromolecule. Mo5-2x/6J1 + G (d,p) density functional theory calculations led to a gas phase stabilization energy of 13.4 kcal/mole for a cofacial complex with one strong and one weak intermolecular bond, versus that of 8.6 kcal/mole due to dynamic electron correlations in the interacting π -constituents alone. These researchers also suggested that head-to-tail orientations of the interacting species were preferred, and additionally proposed that an antiparallel double-stranded lignin template was responsible for obtaining macromolecular lignin domains lacking either crystallinity or optical activity. According to these researchers, the intermolecular interactions between monolignol radicals and the substructures in lignins may be considerably stronger than Watson-Crick A-T or G-C nucleobase pairs in DNA: in the latter, a weak hydrogen bond in an oligonucleotide has been estimated to only amount to an increment of ~ 0.4 – 2.0 kcal/mole to the stabilization energy. While the gas-phase calculations favor the concept of template polymerization, the potential ramifications of this clearly need to be more fully investigated. It will thus be most instructive to establish how many lignin-forming templates there possibly are, and to what extent each would be able to display limited substrate degeneracy. In addition to the question of proposed template replication, another important related emphasis is to unambiguously define the precise molecular basis for the strong associative forces observed between adjacent lignin chains/lignin preparations.
- 6 *Transcriptional Control of Individual Cell Wall Formation Processes, Biomechanics, and Biodegradation of Plant Cell Walls*: With the recent demonstration of specific transcription factors required for fiber formation, this now offers the opportunity to identify how the various cell wall types are differentially generated. This, in turn, may also permit investigation of lignification in specific cell types, and thus as to how they are individually/differentially formed. In any event, this gives another new direction to the possibility of systematically modulating overall plant structure and plant properties for humanity's use – one cell type at a time.

The areas of biomechanics, biodegradation, and factors affecting disease resistance are other most important emphases that need to be enhanced in terms of current levels of scientific investigation, i.e., in order to understand fully the potential of lignin modification. For example, to what extent can lignin contents and composition be manipulated without adversely affecting cell wall properties/vascular integrity for growth/development, harvesting, storage, and further processing? Furthermore, what effects do such manipulations have on biodegradation feasibility (e.g., with different lignin compositions/contents), as well as on plant defense.

In short, there is much still to do in determining how Nature's second most abundant vascular plant biopolymers are formed and their potential (through manipulation) for

humanity's varied needs. The next 5–10 years should be exciting ones for both lignin and plant cell wall research, and promise to be an exciting challenge for the twenty-first century.

Acknowledgments

The authors thank the National Science Foundation (MCB-0417291), the United States Department of Energy (DE FG03-97ER20259), the National Institute of General Medical Sciences (5 R01 GM066173-02), McIntire Stennis, and the G. Thomas and Anita Hargrove Center for Plant Genomic Research for generous financial support.

References

1. Graham, L.E. (1993) *Origin of Land Plants*. John Wiley & Sons, New York.
2. Kenrick, P. & Crane, P.R. (1997) The origin and early evolution of plants on land. *Nature*, **389**, 33–39.
3. Edwards, D. & Wellman, C.H. (2001) Embryophytes on land: The Ordovician to Lochkovian (Lower Devonian) record. In: *Plants Invade the Land: Evolutionary and Environmental Perspectives* (eds. P.G. Gensel & D. Edwards), pp. 3–28. Columbia University Press, New York.
4. Wilson, E.O. (1992) *The Diversity of Life*. W.W. Norton & Company, New York.
5. Sarkanen, K.V. & Hergert, H.L. (1971) Classification and distribution. In: *Lignins. Occurrence, Formation, Structure and Reactions* (eds. K.V. Sarkanen & C.H. Ludwig), pp. 43–94. Wiley-Interscience, New York.
6. Lewis, N.G. & Yamamoto, E. (1989) Tannins—their place in plant metabolism. In: *Chemistry and Significance of Condensed Tannins* (eds. R.W. Hemingway & J.J. Karchesy), pp. 23–46. Plenum Press, New York.
7. Lai, Y.Z. & Sarkanen, K.V. (1971) Isolation and structural studies. In: *Lignins. Occurrence, Formation, Structure and Reactions* (eds. K.V. Sarkanen & C.H. Ludwig), pp. 165–240. Wiley-Interscience, New York.
8. Sarkanen, S., Teller, D.C., Stevens, C.R. & McCarthy, J.L. (1984) Lignin. 20. Associative interactions between kraft lignin components. *Macromolecules*, **17**, 2588–2597.
9. Dutta, S., Garver, T.M., Jr. & Sarkanen, S. (1989) Modes of association between kraft lignin components. In: *Lignin. Properties and Materials*, Vol. 397 (eds. W.G. Glasser & S. Sarkanen), pp. 155–176. ACS Symposium Series, Washington, DC.
10. Erdtman, H. (1933) Dehydrierungen in der Coniferylreihe. II. Dehydrodi-isoeugenol. *Justus Liebigs Annalen der Chemie*, **503**, 283–294.
11. Freudenberg, K. (1965) Lignin: Its constitution and formation from *p*-hydroxycinnamyl alcohols. *Science*, **148**, 595–600.
12. Freudenberg, K. (1968) The constitution and biosynthesis of lignin. In: *Constitution and Biosynthesis of Lignin* (eds. K. Freudenberg & A.C. Neish), pp. 45–122. Springer-Verlag, Berlin.
13. Glasser, W.G. & Glasser, H.R. (1974) Simulation of reactions with lignin by computer (Simrel). II. A model for softwood lignin. *Holzforschung*, **28**, 5–11.
14. Hall, P.L., Glasser, W.G. & Drew, S.W. (1980) Enzymatic transformations of lignin. In: *Lignin Biodegradation: Microbiology, Chemistry, and Potential Applications*, Vol. 2 (eds. T.K. Kirk, T. Higuchi & H.-m. Chang), pp. 33–49. CRC Press, Boca Raton.
15. Brunow, G., Kilpeläinen, I., Sipilä, J., Syrjänen, K., Karhunen, P., Setälä, H. & Rummakko, P. (1998) Oxidative coupling of phenols and the biosynthesis of lignin. In: *Lignin and Lignan Biosynthesis*, Vol. 697 (eds. N.G. Lewis & S. Sarkanen), pp. 131–147. ACS Symposium Series, Washington, DC.

16. Tien, M. & Kirk, T.K. (1983) Lignin-degrading enzyme from the hymenomycete *Phanerochaete chrysosporium* Burds. *Science*, **221**, 661–663.
17. Glenn, J.K., Morgan, M.A., Mayfield, M.B., Kuwahara, M. & Gold, M.H. (1983) An extracellular H₂O₂-requiring enzyme preparation involved in lignin biodegradation by the white rot basidiomycete *Phanerochaete chrysosporium*. *Biochemical and Biophysical Research Communications*, **114**, 1077–1083.
18. Kirk, T.K., Tien, M., Kersten, P.J., Mozuch, M.D. & Kalyanaraman, B. (1986) Ligninase of *Phanerochaete chrysosporium*. Mechanism of its degradation of the non-phenolic arylglycerol β -aryl ether substructure of lignin. *Biochemical Journal*, **236**, 279–287.
19. Tien, M. & Tu, C.-P.D. (1987) Cloning and sequencing of a cDNA for a ligninase from *Phanerochaete chrysosporium*. *Nature*, **326**, 520–523.
20. Call, H.P. & Mücke, I. (1997) History, overview and applications of mediated lignolytic systems, especially laccase-mediator-systems (Lignozym[®]-process). *Journal of Biotechnology*, **53**, 163–202.
21. Chen, Y.-r., Jacobson, B., Sarkanen, S. & Wang, Y. (2001) The preliminary characterization of lignin depolymerase. *Proceedings of the 11th International Symposium on Wood and Pulp Chemistry*, **1**, 313–316.
22. Raven, J.A. (1993) The evolution of vascular plants in relation to quantitative functioning of dead water-conducting cells and stomata. *Biological Reviews*, **68**, 337–363.
23. Kenrick, P. & Crane, P.R. (1997) *The Origin and Early Diversification of Land Plants: A Cladistic Study*. Smithsonian Institution Press, Washington, DC.
24. Bolwell, G.P., Patten, A.M. & Lewis, N.G. (2001) The Holy Grail of wood evolution—from wood anatomy to tissue-specific gene expression: To what extent do molecular studies of biosynthesis of cell wall biopolymers help the understanding of the evolution of woody species? *Phytochemistry*, **57**, 805–810.
25. Bierhorst, D.W. (1960) Observations on tracheary elements. *Phytomorphology*, **10**, 249–305.
26. Gifford, E.M. & Foster, A.S. (1989) *Morphology and Evolution of Vascular Plants*. W.H. Freeman and Company, New York.
27. Ogura, Y. (1972) *Comparative Anatomy of Vegetative Organs of the Pteridophytes*, 2nd edn. Gebrüder Bornträger, Berlin.
28. Bailey, I.W. & Tupper, W.W. (1918) Size variation in tracheary cells: A comparison between the secondary xylems of vascular cryptogams, gymnosperms, and angiosperms. *Proceedings of the American Academy of Arts and Sciences*, **54**, 149–204.
29. Davin, L.B. & Lewis, N.G. (2005) Lignin primary structures and dirigent sites. *Current Opinion in Biotechnology*, **16**, 407–415.
30. Lewis, N.G. & Yamamoto, E. (1990) Lignin: Occurrence, biogenesis and biodegradation. *Annual Review of Plant Physiology and Plant Molecular Biology*, **41**, 455–496.
31. Lewis, N.G., Davin, L.B. & Sarkanen, S. (1999) The nature and function of lignins. In: *Comprehensive Natural Products Chemistry*, Vol. 3 (eds. Sir D.H.R. Barton, K. Nakanishi & O. Meth-Cohn), pp. 617–745. Elsevier, Oxford.
32. Sarkanen, K.V. & Ludwig, C.H. (1971) *Lignins. Occurrence, Formation, Structure and Reactions*. Wiley-Interscience, New York.
33. Sarkanen, K.V. & Ludwig, C.H. (1971) Definition and nomenclature. In: *Lignins. Occurrence, Formation, Structure and Reactions* (eds. K.V. Sarkanen & C.H. Ludwig), pp. 1–18. Wiley-Interscience, New York.
34. Anterola, A.M., van Rensburg, H., van Heerden, P.S., Davin, L.B. & Lewis, N.G. (1999) Multi-site modulation of flux during monolignol formation in loblolly pine (*Pinus taeda*). *Biochemical and Biophysical Research Communications*, **261**, 652–657.
35. Anterola, A.M., Jeon, J.-H., Davin, L.B. & Lewis, N.G. (2002) Transcriptional control of monolignol biosynthesis in *Pinus taeda*: Factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. *Journal of Biological Chemistry*, **277**, 18272–18280.

36. Fergus, B.J. & Goring, D.A.I. (1970) The distribution of lignin in birch wood as determined by ultraviolet microscopy. *Holzforschung*, **24**, 118–124.
37. Towers, G.H.N. & Gibbs, R.D. (1953) Lignin chemistry and the taxonomy of higher plants. *Nature*, **172**, 25–26.
38. White, E. & Towers, G.H.N. (1967) Comparative biochemistry of the lycopods. *Phytochemistry*, **6**, 663–667.
39. Jin, Z., Matsumoto, Y., Tange, T., Akiyama, T., Higuchi, M., Ishii, T. & Iiyama, K. (2005) Proof of the presence of guaiacyl-syringyl lignin in *Selaginella tamariscina*. *Journal of Wood Science*, **51**, 424–426.
40. Logan, K.J. & Thomas, B.A. (1985) Distribution of lignin derivatives in plants. *New Phytologist*, **99**, 571–585.
41. Gibbs, R.D. (1958) The Mäule reaction, lignins, and the relationships between woody plants. In: *The Physiology of Forest Trees* (eds. K.V. Thimann, W.B. Critchfield & M.H. Zimmermann), pp. 269–312. Ronald Press Company, New York.
42. Fergus, B.J., Procter, A.R., Scott, J.A.N. & Goring, D.A.I. (1969) The distribution of lignin in sprucewood as determined by ultraviolet microscopy. *Wood Science and Technology*, **3**, 117–138.
43. Wardrop, A.B. (1971) Occurrence and formation in plants. In: *Lignins. Occurrence, Formation, Structure and Reactions* (eds. K.V. Sarkanen & C.H. Ludwig), pp. 19–41. Wiley-Interscience, New York.
44. Whiting, P. & Goring, D.A.I. (1982) Chemical characterization of tissue fractions from the middle lamella and secondary wall of black spruce tracheids. *Wood Science and Technology*, **16**, 261–267.
45. Terashima, N., Fukushima, K. & Tsuchiya, S. (1986) Heterogeneity in formation of lignin. VII. An autoradiographic study on the formation of guaiacyl and syringyl lignin in poplar. *Journal of Wood Chemistry and Technology*, **6**, 495–504.
46. Terashima, N. & Fukushima, K. (1988) Heterogeneity in formation of lignin. XI. An autoradiographic study of the heterogeneous formation and structure of pine lignin. *Wood Science and Technology*, **22**, 259–270.
47. Fukushima, K. & Terashima, N. (1991) Heterogeneity in formation of lignin. Part XV. Formation and structure of lignin in compression wood of *Pinus thunbergii* studied by microautoradiography. *Wood Science and Technology*, **25**, 371–381.
48. Donaldson, L.A. (1994) Mechanical constraints on lignin deposition during lignification. *Wood Science and Technology*, **28**, 111–118.
49. Donaldson, L.A. (2001) Lignification and lignin topochemistry – an ultrastructural view. *Phytochemistry*, **57**, 859–873.
50. Guan, S.-Y., Mlynár, J. & Sarkanen, S. (1997) Dehydrogenative polymerization of coniferyl alcohol on macromolecular lignin templates. *Phytochemistry*, **45**, 911–918.
51. Sarkanen, S. (1998) Template polymerization in lignin biosynthesis. In: *Lignin and Lignan Biosynthesis*, Vol. 697 (eds. N.G. Lewis & S. Sarkanen), pp. 194–208. ACS Symposium Series, Washington, DC.
52. Chen, Y.-r. & Sarkanen, S. (2003) Macromolecular lignin replication: A mechanistic working hypothesis. *Phytochemistry Reviews*, **2**, 235–255.
53. Freudenberg, K. (1959) Biochimie et constitution de la lignine. *Bulletin de la Société Chimique de France*, 1748–1753.
54. Freudenberg, K. (1959) Biosynthesis and constitution of lignin. *Nature*, **183**, 1152–1155.
55. Kim, S.-J., Kim, K.-W., Cho, M.-H., Franceschi, V.R., Davin, L.B. & Lewis, N.G. (2007) Expression of cinnamyl alcohol dehydrogenases and their putative homologues during *Arabidopsis thaliana* growth and development: Lessons for database annotations? *Phytochemistry*, **68**, 1957–1974.
56. Kim, S.-J., Kim, M.-R., Bedgar, D.L., Moinuddin, S.G.A., Cardenas, C.L., Davin, L.B., Kang, C. & Lewis, N.G. (2004) Functional reclassification of the putative cinnamyl alcohol dehydrogenase

- multigene family in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA*, **101**, 1455–1460.
57. Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapierre, C., Jouanin, L. & Séguin, A. (2005) *CIN-NAMYL ALCOHOL DEHYDROGENASE-C* and *-D* are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. *Plant Cell*, **17**, 2059–2076.
 58. Jorgenson, L.R. (1931) Brown midrib in maize and its linkage relations. *Journal of the American Society of Agronomy*, **23**, 549–557.
 59. Burnham, C.R. & Brink, R.A. (1932) Linkage relations of a second brown midrib gene (*bm*₂) in maize. *Journal of the American Society of Agronomy*, **24**, 960–963.
 60. Emerson, R.A., Beadle, G.W. & Fraser, A.C. (1935) A summary of linkage studies in maize. *Cornell University Agricultural Experiment Station Memoir*, **180**, 1–83.
 61. Kuc, J. & Nelson, O.E. (1964) The abnormal lignins produced by the brown-midrib mutants of maize. I. The brown-midrib-1 mutant. *Archives of Biochemistry and Biophysics*, **105**, 103–113.
 62. Kuc, J., Nelson, O.E. & Flanagan, P. (1968) Degradation of abnormal lignins in the brown-midrib mutants and double mutants of maize. *Phytochemistry*, **7**, 1435–1436.
 63. Gordon, A.J. & Neudoerffer, T.S. (1973) Chemical and *in vivo* evaluation of a brown midrib mutant of *Zea mays*. I. Fibre, lignin, and amino acid composition and digestibility for sheep. *Journal of the Science of Food and Agriculture*, **24**, 565–577.
 64. Hartley, R.D. & Jones, E.C. (1978) Phenolic components and degradability of the cell walls of the brown midrib mutant, *bm*₃, of *Zea mays*. *Journal of the Science of Food and Agriculture*, **29**, 777–789.
 65. Lapierre, C., Tollier, M.-T. & Monties, B. (1988) Mise en évidence d'un nouveau type d'unité constitutive dans les lignines d'un mutant de maïs *bm*₃. *Comptes Rendus de l'Académie des Sciences, Paris, Série III*, **307**, 723–728.
 66. Vignols, F., Rigau, J., Torres, M.A., Capellades, M. & Puigdomènech, P. (1995) The brown midrib3 (*bm*₃) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. *Plant Cell*, **7**, 407–416.
 67. Halpin, C., Holt, K., Chojecki, J., Oliver, D., Chabbert, B., Monties, B., Edwards, K., Barakate, A. & Foxon, G.A. (1998) Brown-midrib maize (*bm*₁) – a mutation affecting the cinnamyl alcohol dehydrogenase gene. *Plant Journal*, **14**, 545–553.
 68. Hepworth, D.G. & Vincent, J.F.V. (1998) The mechanical properties of xylem tissue from tobacco plants (*Nicotiana tabacum* “Samsun”). *Annals of Botany*, **81**, 751–759.
 69. Hepworth, D.G., Vincent, J.F.V. & Schuch, W. (1998) Using viscoelastic properties of the woody tissue from tobacco plants (*Nicotiana tabacum*) to comment on the molecular structure of cell walls. *Annals of Botany*, **81**, 729–734.
 70. Hepworth, D.G. & Vincent, J.F.V. (1999) The growth response of the stems of genetically modified tobacco plants (*Nicotiana tabacum* “Samsun”) to flexural stimulation. *Annals of Botany*, **83**, 39–43.
 71. Jourdes, M., Cardenas, C.L., Laskar, D.D., Moinuddin, S.G.A., Davin, L.B. & Lewis, N.G. (2007) Plant cell walls are enfeebled when attempting to preserve native lignin configuration with poly-*p*-hydroxycinnamaldehydes: Evolutionary implications. *Phytochemistry*, **68**, 1932–1956.
 72. Patten, A.M., Jourdes, M., Brown, E.E., Laborie, M.-P., Davin, L.B. & Lewis, N.G. (2007) Reaction tissue formation and stem tensile modulus properties in wild-type and *p*-coumarate-3-hydroxylase downregulated lines of alfalfa, *Medicago sativa* (Fabaceae). *American Journal of Botany*, **94**, 912–925.
 73. Chattaway, M.M. (1949) The development of tyloses and secretion of gum in heartwood formation. *Australian Journal of Scientific Research*, **2**, 227–240.
 74. Gang, D.R., Fujita, M., Davin, L.B. & Lewis, N.G. (1998) The “abnormal lignins”: Mapping heartwood formation through the lignan biosynthetic pathway. In: *Lignin and Lignan Biosynthesis*, Vol. 697 (eds N.G. Lewis & S. Sarkanen), pp. 389–421. ACS Symposium Series, Washington, DC.

75. Kwon, M., Bedgar, D.L., Piastuch, W., Davin, L.B. & Lewis, N.G. (2001) Induced compression wood formation in Douglas fir (*Pseudotsuga menziesii*) in microgravity. *Phytochemistry*, **57**, 847–857.
76. Wardrop, A.B. & Davies, G.W. (1964) The nature of reaction wood. VIII. The structure and differentiation of compression wood. *Australian Journal of Botany*, **12**, 24–38.
77. Anterola, A.M. & Lewis, N.G. (2002) Trends in lignin modification: A comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry*, **61**, 221–294.
78. Humphreys, J.M. & Chapple, C. (2002) Rewriting the lignin roadmap. *Current Opinion in Plant Biology*, **5**, 224–229.
79. The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.
80. Cotton, R.G.H. & Gibson, F. (1968) The biosynthesis of phenylalanine and tyrosine in the pea (*Pisum sativum*): Chorismate mutase. *Biochimica et Biophysica Acta*, **156**, 187–189.
81. Jung, E., Zamir, L.O. & Jensen, R.A. (1986) Chloroplasts of higher plants synthesize L-phenylalanine via L-arogenate. *Proceedings of the National Academy of Sciences, USA*, **83**, 7231–7235.
82. Siehl, D.L. & Conn, E.E. (1988) Kinetic and regulatory properties of arogenate dehydratase in seedlings of *Sorghum bicolor* (L.) Moench. *Archives of Biochemistry and Biophysics*, **260**, 822–829.
83. Cho, M.-H., Corea, O.R.A., Yang, H., Bedgar, D.L., Laskar, D.D., Anterola, A.M., Moog-Anterola, F.A., Hood, R.L., Kohalmi, S.E., Bernards, M.A., Kang, C., Davin, L.B. & Lewis, N.G. (2007) Phenylalanine biosynthesis in *Arabidopsis thaliana*: Identification and characterization of arogenate dehydratases. *Journal of Biological Chemistry*, **282**, 30827–30835.
84. Koukol, J. & Conn, E.E. (1961) The metabolism of aromatic compounds in higher plants. IV. Purification and properties of the phenylalanine deaminase of *Hordeum vulgare*. *Journal of Biological Chemistry*, **236**, 2692–2698.
85. Razal, R.A., Ellis, S., Singh, S., Lewis, N.G. & Towers, G.H.N. (1996) Nitrogen recycling in phenylpropanoid metabolism. *Phytochemistry*, **41**, 31–35.
86. van Heerden, P.S., Towers, G.H.N. & Lewis, N.G. (1996) Nitrogen metabolism in lignifying *Pinus taeda* cell cultures. *Journal of Biological Chemistry*, **271**, 12350–12355.
87. Singh, S., Lewis, N.G. & Towers, G.H.N. (1997) Interaction between nitrogen and phenylpropanoid metabolism: Detection of a novel nitrogen cycle in plants. *Polyphénols Actualités*, **15/16**, 16–19.
88. Singh, S., Lewis, N.G. & Towers, G.H.N. (1998) Nitrogen recycling during phenylpropanoid metabolism in sweet potato tubers. *Journal of Plant Physiology*, **153**, 316–323.
89. Edwards, K., Cramer, C.L., Bolwell, G.P., Dixon, R.A., Schuch, W. & Lamb, C.J. (1985) Rapid transient induction of phenylalanine ammonia-lyase mRNA in elicitor-treated bean cells. *Proceedings of the National Academy of Sciences, USA*, **82**, 6731–6735.
90. Kuhn, D.N., Chappell, J. & Hahlbrock, K. (1983) Identification and use of cDNAs of phenylalanine ammonia-lyase and 4-coumarate:CoA ligase mRNAs in studies of the induction of phytoalexin biosynthetic enzymes in cultured parsley cells. In: *Structure and Function of Plant Genes* (eds. O. Ciferri & L. Dure), pp. 329–336. Chapman and Hall, London.
91. Kuhn, D.N., Chappell, J., Boudet, A. & Hahlbrock, K. (1984) Induction of phenylalanine ammonia-lyase and 4-coumarate:CoA ligase mRNAs in cultured plant cells by UV light or fungal elicitor. *Proceedings of the National Academy of Sciences, USA*, **81**, 1102–1106.
92. Lois, R., Dietrich, A., Hahlbrock, K. & Schulz, W. (1989) A phenylalanine ammonia-lyase gene from parsley: Structure, regulation and identification of elicitor and light responsive *cis*-acting elements. *EMBO Journal*, **8**, 1641–1648.
93. Costa, M.A., Collins, R.E., Anterola, A.M., Cochrane, F.C., Davin, L.B. & Lewis, N.G. (2003) An *in silico* assessment of gene function and organization of the phenylpropanoid pathway

- metabolic networks in *Arabidopsis thaliana* and limitations thereof. *Phytochemistry*, **64**, 1097–1112.
94. Cochrane, F.C., Davin, L.B. & Lewis, N.G. (2004) The *Arabidopsis* phenylalanine ammonia-lyase gene family: Kinetic characterization of the four PAL isoforms. *Phytochemistry*, **65**, 1557–1564.
 95. Calabrese, J.C., Jordan, D.B., Boodhoo, A., Sariaslani, S. & Vannelli, T. (2004) Crystal structure of phenylalanine ammonia lyase: Multiple helix dipoles implicated in catalysis. *Biochemistry*, **43**, 11403–11416.
 96. Schwede, T.F., Rétey, J. & Schulz, G.E. (1999) Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry*, **38**, 5355–5361.
 97. Baedeker, M. & Schulz, G.E. (2002) Autocatalytic peptide cyclization during chain folding of histidine ammonia-lyase. *Structure*, **10**, 61–67.
 98. Watts, K.T., Mijts, B.N., Lee, P.C., Manning, A.J. & Schmidt-Dannert, C. (2006) Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. *Chemistry and Biology*, **13**, 1317–1326.
 99. Louie, G.V., Bowman, M.E., Moffitt, M.C., Baiga, T.J., Moore, B.S. & Noel, J.P. (2006) Structural determinants and modulation of substrate specificity in phenylalanine-tyrosine ammonia-lyases. *Chemistry and Biology*, **13**, 1327–1338.
 100. Weng, J.-K. & Chapple, C. (2007) The lycophyte *Selaginella moellendorffii*: An emerging plant model for studying comparative genomics and the evolution of phenylpropanoid metabolism [Abstract: W247]. In: *Plant and Animal Genomes XV Conference*, San Diego, California, January 13–17.
 101. Russell, D.W. & Conn, E.E. (1967) The cinnamic acid 4-hydroxylase of pea seedlings. *Archives of Biochemistry and Biophysics*, **122**, 256–258.
 102. Teutsch, H.G., Hasenfratz, M.P., Lesot, A., Stoltz, C., Garnier, J.-M., Jeltsch, J.-M., Durst, F. & Werck-Reichhart, D. (1993) Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. *Proceedings of the National Academy of Sciences, USA*, **90**, 4102–4106.
 103. Mizutani, M., Ward, E., DiMaio, J., Ohta, D., Ryals, J. & Sato, R. (1993) Molecular cloning and sequencing of a cDNA encoding mung bean cytochrome P450 (P450C4H) possessing cinnamate 4-hydroxylase activity. *Biochemical and Biophysical Research Communications*, **190**, 875–880.
 104. Vaughan, P.F.T. & Butt, V.S. (1969) The hydroxylation of *p*-coumaric acid by an enzyme from leaves of spinach beet (*Beta vulgaris* L.). *Biochemical Journal*, **113**, 109–115.
 105. Kojima, M. & Takeuchi, W. (1989) Detection and characterization of *p*-coumaric acid hydroxylase in mung bean, *Vigna mungo*, seedlings. *Journal of Biochemistry*, **105**, 265–270.
 106. Kamsteeg, J., van Brederode, J., Verschuren, P.M. & van Nigtevecht, G. (1981) Identification, properties and genetic control of *p*-coumaroyl-coenzyme A 3-hydroxylase isolated from petals of *Silene dioica*. *Zeitschrift für Pflanzenphysiologie*, **102**, 435–442.
 107. Heller, W. & Kühnl, T. (1985) Elicitor induction of a microsomal 5-O-(4-coumaroyl)shikimate 3'-hydroxylase in parsley cell suspension cultures. *Archives of Biochemistry and Biophysics*, **241**, 453–460.
 108. Schoch, G., Goepfert, S., Morant, M., Hehn, A., Meyer, D., Ullmann, P. & Werck-Reichhart, D. (2001) CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *Journal of Biological Chemistry*, **276**, 36566–36574.
 109. Stöckigt, J. & Zenk, M.H. (1974) Enzymatic synthesis of chlorogenic acid from caffeoyl coenzyme A and quinic acid. *FEBS Letters*, **42**, 131–134.
 110. Hoffmann, L., Maury, S., Martz, F., Geoffroy, P. & Legrand, M. (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *Journal of Biological Chemistry*, **278**, 95–103.
 111. Niggeweg, R., Michael, A.J. & Martin, C. (2004) Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nature Biotechnology*, **22**, 746–754.

112. Grand, C. (1984) Ferulic acid 5-hydroxylase: A new cytochrome P-450-dependent enzyme from higher plant microsomes involved in lignin synthesis. *FEBS Letters*, **169**, 7–11.
113. Ohashi, H., Yamamoto, E., Lewis, N.G. & Towers, G.H.N. (1987) 5-Hydroxyferulic acid in *Zea mays* and *Hordeum vulgare* cell walls. *Phytochemistry*, **26**, 1915–1916.
114. Chen, F., Yasuda, S. & Fukushima, K. (1999) Evidence for a novel biosynthetic pathway that regulates the ratio of syringyl to guaiacyl residues in lignin in the differentiating xylem of *Magnolia kobus* DC. *Planta*, **207**, 597–603.
115. Chen, F., Fukushima, K. & Yasuda, S. (1999) The actual route in the biosynthesis of lignin precursors. *10th International Symposium on Wood and Pulp Chemistry*, **1**, 134–137.
116. Humphreys, J.M., Hemm, M.R. & Chapple, C. (1999) New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. *Proceedings of the National Academy of Sciences, USA*, **96**, 10045–10050.
117. Gross, G.G. & Zenk, M.H. (1966) Darstellung und Eigenschaften von Coenzym A-Thiolesteren substituierter Zimtsäuren. *Zeitschrift für Naturforschung*, **21b**, 683–690.
118. Gross, G.G., Stöckigt, J., Mansell, R.L. & Zenk, M.H. (1973) Three novel enzymes involved in the reduction of ferulic acid to coniferyl alcohol in higher plants: Ferulate:CoA ligase, feruloyl-CoA reductase and coniferyl alcohol oxidoreductase. *FEBS Letters*, **31**, 283–286.
119. Hahlbrock, K. & Grisebach, H. (1970) Formation of coenzyme A esters of cinnamic acids with an enzyme preparation from cell suspension cultures of parsley. *FEBS Letters*, **11**, 62–64.
120. Mansell, R.L., Stöckigt, J. & Zenk, M.H. (1972) Reduction of ferulic acid to coniferyl alcohol in a cell free system from a higher plant. *Zeitschrift für Pflanzenphysiologie*, **68**, 286–288.
121. Douglas, C., Hoffmann, H., Schulz, W. & Hahlbrock, K. (1987) Structure and elicitor or U.V.-light-stimulated expression of two 4-coumarate:CoA ligase genes in parsley. *EMBO Journal*, **6**, 1189–1195.
122. Costa, M.A., Bedgar, D.L., Moinuddin, S.G.A., Kim, K.-W., Cardenas, C.L., Cochrane, F.C., Shockey, J.M., Helms, G.L., Amakura, Y., Takahashi, H., Milhollan, J.K., Davin, L.B., Browse, J. & Lewis, N.G. (2005) Characterization *in vitro* and *in vivo* of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: Syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry*, **66**, 2072–2091.
123. Mansell, R.L., Gross, G.G., Stöckigt, J., Franke, H. & Zenk, M.H. (1974) Purification and properties of cinnamyl alcohol dehydrogenase from higher plants involved in lignin biosynthesis. *Phytochemistry*, **13**, 2427–2435.
124. Gross, G.G. & Kreiten, W. (1975) Reduction of coenzyme A thioesters of cinnamic acids with an enzyme preparation from lignifying tissue of *Forsythia*. *FEBS Letters*, **54**, 259–262.
125. Lacombe, E., Hawkins, S., Van Doorselaere, J., Piquemal, J., Goffner, D., Poeydomenge, O., Boudet, A.M. & Grima-Pettenati, J. (1997) Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: Cloning, expression and phylogenetic relationships. *Plant Journal*, **11**, 429–441.
126. Knight, M.E., Halpin, C. & Schuch, W. (1992) Identification and characterization of cDNA clones encoding cinnamyl alcohol dehydrogenase from tobacco. *Plant Molecular Biology*, **19**, 793–801.
127. Walter, M.H., Grima-Pettenati, J., Grand, C., Boudet, A.M. & Lamb, C.J. (1988) Cinnamyl alcohol dehydrogenase, a molecular marker specific for lignin synthesis: cDNA cloning and mRNA induction by fungal elicitor. *Proceedings of the National Academy of Sciences, USA*, **85**, 5546–5550.
128. Walter, M.H., Grima-Pettenati, J., Grand, C., Boudet, A.M. & Lamb, C.J. (1990) Extensive sequence similarity of the bean CAD4 (cinnamyl alcohol dehydrogenase) to a maize malic enzyme. *Plant Molecular Biology*, **15**, 525–526.
129. Walter, M.H., Grima-Pettenati, J. & Feuillet, C. (1994) Characterization of a bean (*Phaseolus vulgaris* L.) malic-enzyme gene. *European Journal of Biochemistry*, **224**, 999–1009.

130. Lauvergeat, V., Lacomme, C., Lacombe, E., Lasserre, E., Roby, D. & Grima-Pettenati, J. (2001) Two cinnamoyl-CoA reductase (CCR) genes from *Arabidopsis thaliana* are differentially expressed during development and in response to infection with pathogenic bacteria. *Phytochemistry*, **57**, 1187–1195.
131. Patten, A.M., Cardenas, C.L., Cochrane, F.C., Laskar, D.D., Bedgar, D.L., Davin, L.B. & Lewis, N.G. (2005) Reassessment of effects on lignification and vascular development in the *irx4* *Arabidopsis* mutant. *Phytochemistry*, **66**, 2092–2107.
132. Laskar, D.D., Jourdes, M., Patten, A.M., Helms, G.L., Davin, L.B. & Lewis, N.G. (2006) The *Arabidopsis* cinnamoyl CoA reductase *irx4* mutant has a delayed but coherent (normal) program of lignification. *Plant Journal*, **48**, 674–686.
133. Youn, B., Camacho, R., Moinuddin, S.G.A., Lee, C., Davin, L.B., Lewis, N.G. & Kang, C. (2006) Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Organic and Biomolecular Chemistry*, **4**, 1687–1697.
134. Finkle, B.J. & Nelson, R.F. (1963) Enzyme reactions with phenolic compounds: A *meta*-O-methyltransferase in plants. *Biochimica et Biophysica Acta*, **78**, 747–749.
135. Pakusch, A.-E., Kneusel, R.E. & Matern, U. (1989) S-Adenosyl-L-methionine:trans-caffeoyl-coenzyme A 3-O-methyltransferase from elicitor-treated parsley cell suspension cultures. *Archives of Biochemistry and Biophysics*, **271**, 488–494.
136. Atanassova, R., Favet, N., Martz, F., Chabbert, B., Tollier, M.-T., Monties, B., Fritig, B. & Legrand, M. (1995) Altered lignin composition in transgenic tobacco expressing O-methyltransferase sequences in sense and antisense orientation. *Plant Journal*, **8**, 465–477.
137. Bugos, R.C., Chiang, V.L.C. & Campbell, W.H. (1991) cDNA cloning, sequence analysis and seasonal expression of lignin-bispecific caffeic acid/5-hydroxyferulic acid O-methyltransferase of aspen. *Plant Molecular Biology*, **17**, 1203–1215.
138. Gowri, G., Bugos, R.C., Campbell, W.H., Maxwell, C.A. & Dixon, R.A. (1991) Stress responses in alfalfa (*Medicago sativa* L.). X. Molecular cloning and expression of S-adenosyl-L-methionine: caffeic acid 3-O-methyltransferase, a key enzyme of lignin biosynthesis. *Plant Physiology*, **97**, 7–14.
139. Schmitt, D., Pakusch, A.-E. & Matern, U. (1991) Molecular cloning, induction, and taxonomic distribution of caffeoyl-CoA 3-O-methyltransferase, an enzyme involved in disease resistance. *Journal of Biological Chemistry*, **266**, 17416–17423.
140. Goffner, D., Van Doorselaere, J., Yahiaoui, N., Samaj, J., Grima-Pettenati, J. & Boudet, A.M. (1998) A novel aromatic alcohol dehydrogenase in higher plants: Molecular cloning and expression. *Plant Molecular Biology*, **36**, 755–765.
141. Damiani, I., Morreel, K., Danoun, S., Goeminne, G., Yahiaoui, N., Marque, C., Kopka, J., Messens, E., Goffner, D., Boerjan, W., Boudet, A.-M. & Rochange, S. (2005) Metabolite profiling reveals a role for atypical cinnamyl alcohol dehydrogenase CAD1 in the synthesis of coniferyl alcohol in tobacco xylem. *Plant Molecular Biology*, **59**, 753–769.
142. Goffner, D., Joffroy, I., Grima-Pettenati, J., Halpin, C., Knight, M.E., Schuch, W. & Boudet, A.M. (1992) Purification and characterization of isoforms of cinnamyl alcohol dehydrogenase from *Eucalyptus* xylem. *Planta*, **188**, 48–53.
143. Li, L., Cheng, X.F., Leshkevich, J., Umezawa, T., Harding, S.A. & Chiang, V.L. (2001) The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. *Plant Cell*, **13**, 1567–1585.
144. Bomati, E.K. & Noel, J.P. (2005) Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell*, **17**, 1598–1611.
145. Zhang, K., Qian, Q., Huang, Z., Wang, Y., Li, M., Hong, L., Zeng, D., Gu, M., Chu, C. & Cheng, Z. (2006) GOLD HULL AND INTERNODE2 encodes a primarily multifunctional cinnamyl-alcohol dehydrogenase in rice. *Plant Physiology*, **140**, 972–983.

146. Chen, H., Jiang, H. & Morgan, J.A. (2007) Non-natural cinnamic acid derivatives as substrates of cinnamate 4-hydroxylase. *Phytochemistry*, **68**, 306–311.
147. Lewis, N.G. & Yean, W.Q. (1985) High performance size-exclusion chromatography of lignosulphonates. *Journal of Chromatography*, **331**, 419–424.
148. Bialski, A.M., Luthe, C.E., Fong, J.L. & Lewis, N.G. (1986) Sulphite-promoted delignification of wood: Identification of paucidisperse lignosulphonates. *Canadian Journal of Chemistry*, **64**, 1336–1344.
149. Bialski, A.M., Bradford, H., Lewis, N.G. & Luthe, C.E. (1986) Lignosulphonate polymerization: Effect of cross-linking agents. *Journal of Applied Polymer Science*, **31**, 1363–1372.
150. Luthe, C.E. & Lewis, N.G. (1986) Identification and characterization of paucidisperse lignosulphonates. *Holzforschung*, **40**(suppl), 153–157.
151. Björkman, A. (1954) Isolation of lignin from finely divided wood with neutral solvents. *Nature*, **174**, 1057–1058.
152. Björkman, A. (1957) Lignin and lignin-carbohydrate complexes. Extraction from wood meal with neutral solvents. *Industrial and Engineering Chemistry*, **49**, 1395–1398.
153. Lu, F. & Ralph, J. (2003) Non-degradative dissolution and acetylation of ball-milled plant cell walls: High-resolution solution-state NMR. *Plant Journal*, **35**, 535–544.
154. Nimz, H.H. & Lüdemann, H.-D. (1976) Carbon-13 NMR spectra of lignins. 6. Lignin and DHP acetates. *Holzforschung*, **30**, 33–40.
155. Nimz, H.H. & Tutschek, R. (1977) Carbon-13 NMR spectra of lignins. 7. On the question of the lignin content of mosses (*Sphagnum magellanicum* Brid.). *Holzforschung*, **31**, 101–106.
156. Nimz, H.H. (1978) Carbon-13 NMR spectra of lignins. *Bulletin de Liaison – Groupe Polyphénols*, **8**, 185–206.
157. Nimz, H.H., Robert, D., Faix, O. & Nemr, M. (1981) Carbon-13 NMR spectra of lignins. 8. Structural differences between lignins of hardwoods, softwoods, grasses and compression wood. *Holzforschung*, **35**, 16–26.
158. Nimz, H.H., Nemr, M., Schmidt, P., Margot, C., Schaub, B. & Schlosser, M. (1982) Carbon-13 NMR spectra of lignins. 9. Spin-lattice relaxation times (T_1) and determination of interunit linkages in three hardwood lignins (*Alnus glutinosa*, *Corylus avellanus* and *Acer pseudoplatanus*). *Journal of Wood Chemistry and Technology*, **2**, 371–382.
159. Nimz, H.H., Tschirner, U., Stähle, M., Lehmann, R. & Schlosser, M. (1984) Carbon-13 NMR spectra of lignins. 10. Comparison of structural units in spruce and beech lignin. *Journal of Wood Chemistry and Technology*, **4**, 265–284.
160. Mollard, A. & Robert, D. (1984) Etude de la lignine pariétale et extracellulaire des suspensions cellulaires de *Rosa glauca*. *Physiologie Végétale*, **22**, 3–17.
161. Bardet, M., Foray, M.-F. & Robert, D. (1985) Use of the DEPT pulse sequence to facilitate the ^{13}C NMR structural analysis of lignins. *Makromolekulare Chemie*, **186**, 1495–1504.
162. Lewis, N.G., Newman, J., Just, G. & Ripmeister, J. (1987) Determination of bonding patterns of ^{13}C specifically enriched dehydrogenatively polymerized lignin in solution and solid state. *Macromolecules*, **20**, 1752–1756.
163. Lewis, N.G., Yamamoto, E., Wooten, J.B., Just, G., Ohashi, H. & Towers, G.H.N. (1987) Monitoring biosynthesis of wheat cell-wall phenylpropanoids *in situ*. *Science*, **237**, 1344–1346.
164. Lewis, N.G., Razal, R.A., Dhara, K.P., Yamamoto, E., Bokelman, G.H. & Wooten, J.B. (1988) Incorporation of $[2-^{13}\text{C}]$ ferulic acid, a lignin precursor, into *Leucaena leucocephala* and its analysis by solid state ^{13}C NMR spectroscopy. *Journal of the Chemical Society, Chemical Communications*, 1626–1628.
165. Chen, C.-L. & Robert, D. (1988) Characterization of lignin by ^1H and ^{13}C NMR spectroscopy. *Methods in Enzymology*, **161**, 137–174.

166. Robert, D., Mollard, A. & Barnoud, F. (1989) ^{13}C NMR qualitative and quantitative study of lignin structure synthesized in *Rosa glauca* calluses. *Plant Physiology and Biochemistry*, **27**, 297–304.
167. Lewis, N.G., Razal, R.A., Yamamoto, E., Bokelman, G.H. & Wooten, J.B. (1989) ^{13}C Specific labeling of lignin in intact plants. In: *Plant Cell Wall Polymers: Biogenesis and Biodegradation*, Vol. 399 (eds. N.G. Lewis & M.G. Paice), pp. 169–181. ACS Symposium Series, Washington, DC.
168. Terashima, N., Seguchi, Y. & Robert, D. (1991) Selective ^{13}C enrichment of side chain carbons of guaiacyl lignin in pine. *Holzforschung*, **45**(suppl), 35–39.
169. Eberhardt, T.L., Bernards, M.A., He, L., Davin, L.B., Wooten, J.B. & Lewis, N.G. (1993) Lignification in cell suspension cultures of *Pinus taeda*. *In situ* characterization of a gymnosperm lignin. *Journal of Biological Chemistry*, **268**, 21088–21096.
170. Kilpeläinen, I., Ämmälähti, E., Brunow, G. & Robert, D. (1994) Application of three-dimensional HMQC-HOHAHA NMR spectroscopy to wood lignin, a natural polymer. *Tetrahedron Letters*, **35**, 9267–9270.
171. Nose, M., Bernards, M.A., Furlan, M., Zajicek, J., Eberhardt, T.L. & Lewis, N.G. (1995) Towards the specification of consecutive steps in macromolecular lignin assembly. *Phytochemistry*, **39**, 71–79.
172. Ralph, J., MacKay, J.J., Hatfield, R.D., O'Malley, D.M., Whetten, R.W. & Sederoff, R.R. (1997) Abnormal lignin in a loblolly pine mutant. *Science*, **277**, 235–239.
173. Ralph, J., Hatfield, R.D., Piquemal, J., Yahiaoui, N., Pean, M., Lapierre, C. & Boudet, A.M. (1998) NMR characterization of altered lignins extracted from tobacco plants down-regulated for lignification enzymes cinnamyl alcohol dehydrogenase and cinnamoyl-CoA reductase. *Proceedings of the National Academy of Sciences, USA*, **95**, 12803–12808.
174. Chabannes, M., Barakate, A., Lapierre, C., Marita, J.M., Ralph, J., Pean, M., Danoun, S., Halpin, C., Grima-Pettenati, J. & Boudet, A.M. (2001) Strong decrease in lignin content without significant alteration of plant development is induced by simultaneous down-regulation of cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) in tobacco plants. *Plant Journal*, **28**, 257–270.
175. Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P.F., Marita, J.M., Hatfield, R.D., Ralph, S.A., Christensen, J.H. & Boerjan, W. (2004) Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochemistry Reviews*, **3**, 29–60.
176. Bardet, M., Lundquist, K., Parkäs, J., Robert, D. & von Unge, S. (2006) ^{13}C Assignments of the carbon atoms in the aromatic rings of lignin model compounds of the arylglycerol β -aryl ether type. *Magnetic Resonance in Chemistry*, **44**, 976–979.
177. Laskar, D.D., Jourdes, M., Davin, L.B. & Lewis, N.G. (2008) Cinnamyl alcohol dehydrogenase downregulation in tobacco: Reassessment of red lignin. *Phytochemistry*, (in press).
178. Hamcerencu, M., Desbrieres, J., Popa, M., Khoukh, A. & Riess, G. (2007) New unsaturated derivatives of Xanthan gum: Synthesis and characterization. *Polymer*, **48**, 1921–1929.
179. Dence, C.W. (1992) The determination of lignin. In: *Methods in Lignin Chemistry* (eds. S.Y. Lin & C.W. Dence), pp. 33–61. Springer-Verlag, Berlin.
180. Leary, G.J., Newman, R.H. & Morgan, K.R. (1986) A carbon-13 nuclear magnetic resonance study of chemical processes involved in the isolation of Klason lignin. *Holzforschung*, **40**, 267–272.
181. Huntley, S.K., Ellis, D., Gilbert, M., Chapple, C. & Mansfield, S.D. (2003) Significant increases in pulping efficiency in C4H-F5H-transformed poplars: Improved chemical savings and reduced environmental toxins. *Journal of Agricultural and Food Chemistry*, **51**, 6178–6183.
182. Van Doorselaere, J., Baucher, M., Chognot, E., Chabbert, B., Tollier, M.-T., Petit-Conil, M., Leplé, J.-C., Pilate, G., Cornu, D., Monties, B., Van Montagu, M., Inzé, D., Boerjan, W. & Jouanin, L. (1995) A novel lignin in poplar trees with a reduced caffeic acid/5-hydroxyferulic acid O-methyltransferase activity. *Plant Journal*, **8**, 855–864.
183. Lapierre, C., Pollet, B., Petit-Conil, M., Toval, G., Romero, J., Pilate, G., Leplé, J.-C., Boerjan, W., Ferret, V., de Nadaï, V. & Jouanin, L. (1999) Structural alterations of lignins in transgenic poplars

- with depressed cinnamyl alcohol dehydrogenase or caffeic acid *O*-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant Physiology*, **119**, 153–163.
184. Jouanin, L., Goujon, T., de Nadaï, V., Martin, M.-T., Mila, I., Vallet, C., Pollet, B., Yoshinaga, A., Chabbert, B., Petit-Conil, M. & Lapierre, C. (2000) Lignification in transgenic poplars with extremely reduced caffeic acid *O*-methyltransferase activity. *Plant Physiology*, **123**, 1363–1373.
185. Lee, D., Meyer, K., Chapple, C. & Douglas, C.J. (1997) Antisense suppression of 4-coumarate: coenzyme A ligase activity in *Arabidopsis* leads to altered lignin subunit composition. *Plant Cell*, **9**, 1985–1998.
186. Ni, W., Paiva, N.L. & Dixon, R.A. (1994) Reduced lignin in transgenic plants containing a caffeic acid *O*-methyltransferase antisense gene. *Transgenic Research*, **3**, 120–126.
187. Iiyama, K. & Wallis, A.F.A. (1990) Determination of lignin in herbaceous plants by an improved acetyl bromide procedure. *Journal of the Science of Food and Agriculture*, **51**, 145–161.
188. Cardenas, C.L., Jourdes, M., Kim, M.R., Davin, L.B. & Lewis, N.G. (2008) (Manuscript submitted).
189. Schultz, T.P. & Templeton, M.C. (1986) Proposed mechanism for the nitrobenzene oxidation of lignin. *Holzforschung*, **40**, 93–97.
190. Schultz, T.P., Fisher, T.H. & Dershem, S.M. (1987) Role of the *p*-hydroxyl group in the nitrobenzene oxidation of hydroxybenzyl alcohols. *Journal of Organic Chemistry*, **52**, 279–281.
191. Gellerstedt, G. (1992) Chemical degradation methods: Permanganate oxidation. In: *Methods in Lignin Chemistry* (eds. S.Y. Lin & C.W. Dence), pp. 322–333. Springer-Verlag, Berlin.
192. Chang, H.-M. & Allan, G.G. (1971) Oxidation. In: *Lignins. Occurrence, Formation, Structure and Reactions* (eds. K.V. Sarkanen & C.H. Ludwig), pp. 433–485. Wiley-Interscience, New York.
193. Lundquist, K. (1970) Acid degradation of lignin. II. Separation and identification of low molecular weight phenols. *Acta Chemica Scandinavica*, **24**, 889–907.
194. Kishimoto, T., Uraki, Y. & Ubukata, M. (2006) Chemical synthesis of β -*O*-4 type artificial lignin. *Organic and Biomolecular Chemistry*, **4**, 1343–1347.
195. Méchin, V., Baumberger, S., Pollet, B. & Lapierre, C. (2007) Peroxidase activity can dictate the *in vitro* lignin dehydrogenative polymer structure. *Phytochemistry*, **68**, 571–579.
196. Ralph, J. & Rodger, C. (1991) NMR of lignin model trimers or why you will never find crystalline regions in lignin! *6th International Symposium on Wood and Pulp Chemistry*, **1**, 59–64.
197. Ralph, J. (1997) Recent advances in characterizing “non-traditional” lignins. *9th International Symposium on Wood and Pulp Chemistry*, **1**, PL2.1–PL2.7.
198. Zhong, R., Demura, T. & Ye, Z.-H. (2006) SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *Plant Cell*, **18**, 3158–3170.
199. Mitsuda, N., Iwase, A., Yamamoto, H., Yoshida, M., Seki, M., Shinozaki, K. & Ohme-Takagi, M. (2007) NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. *Plant Cell*, **19**, 270–280.
200. Amrhein, N. & Gödeke, K.-H. (1977) α -Aminoxy- β -phenylpropionic acid—a potent inhibitor of L-phenylalanine ammonia-lyase *in vitro* and *in vivo*. *Plant Science Letters*, **8**, 313–317.
201. Amrhein, N. & Gerhardt, J. (1979) Superinduction of phenylalanine ammonia-lyase in gherkin hypocotyls caused by the inhibitor, L- α -aminoxy- β -phenylpropionic acid. *Biochimica et Biophysica Acta*, **583**, 434–442.
202. Amrhein, N. & Holländer, H. (1979) Inhibition of anthocyanin formation in seedlings and flowers by the enantiomers of α -aminoxy- β -phenylpropionic acid and their N-benzoyloxycarbonyl derivatives. *Planta*, **144**, 385–389.
203. Amrhein, N. & Diederich, E. (1980) Turnover of isoflavones in *Cicer arietinum* L. *Naturwissenschaften*, **67**, 40–41.
204. Amrhein, N., Frank, G., Lemm, G. & Luhmann, H.-B. (1983) Inhibition of lignin formation by L- α -aminoxy- β -phenylpropionic acid, an inhibitor of phenylalanine ammonia-lyase. *European Journal of Cell Biology*, **29**, 139–144.

205. Smart, C.C. & Amrhein, N. (1985) The influence of lignification on the development of vascular tissue in *Vigna radiata* L. *Protoplasma*, **124**, 87–95.
206. Elkind, Y., Edwards, R., Mavandad, M., Hedrick, S.A., Ribak, O., Dixon, R.A. & Lamb, C.J. (1990) Abnormal plant development and down-regulation of phenylpropanoid biosynthesis in transgenic tobacco containing a heterologous phenylalanine ammonia-lyase gene. *Proceedings of the National Academy of Sciences, USA*, **87**, 9057–9061.
207. Sewalt, V.J.H., Ni, W., Blount, J.W., Jung, H.G., Masoud, S.A., Howles, P.A., Lamb, C. & Dixon, R.A. (1997) Reduced lignin content and altered lignin composition in transgenic tobacco down-regulated in expression of L-phenylalanine ammonia-lyase or cinnamate 4-hydroxylase. *Plant Physiology*, **115**, 41–50.
208. Rohde, A., Morreel, K., Ralph, J., Goeminne, G., Hostyn, V., De Rycke, R., Kushnir, S., Van Doorselaere, J., Joseleau, J.-P., Vuylsteke, M., Van Driessche, G., Van Beeumen, J., Messens, E. & Boerjan, W. (2004) Molecular phenotyping of the *pal1* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *Plant Cell*, **16**, 2749–2771.
209. Korth, K.L., Blount, J.W., Chen, F., Rasmussen, S., Lamb, C. & Dixon, R.A. (2001) Changes in phenylpropanoid metabolites associated with homology-dependent silencing of phenylalanine ammonia-lyase and its somatic reversion in tobacco. *Physiologia Plantarum*, **111**, 137–143.
210. Blount, J.W., Korth, K.L., Masoud, S.A., Rasmussen, S., Lamb, C. & Dixon, R.A. (2000) Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feed-back loop at the entry point into the phenylpropanoid pathway. *Plant Physiology*, **122**, 107–116.
211. Blee, K., Choi, J.W., O'Connell, A.P., Jupe, S.C., Schuch, W., Lewis, N.G. & Bolwell, G.P. (2001) Antisense and sense expression of cDNA coding for CYP73A15, a class II cinnamate 4-hydroxylase, leads to a delayed and reduced production of lignin in tobacco. *Phytochemistry*, **57**, 1159–1166.
212. Sewalt, V.J.H., Ni, W., Jung, H.G. & Dixon, R.A. (1997) Lignin impact on fiber degradation: Increased enzymatic digestibility of genetically engineered tobacco (*Nicotiana tabacum*) stems reduced in lignin content. *Journal of Agricultural and Food Chemistry*, **45**, 1977–1983.
213. Franke, R., Hemm, M.R., Denault, J.W., Ruegger, M.O., Humphreys, J.M. & Chapple, C. (2002) Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of *Arabidopsis*. *Plant Journal*, **30**, 47–59.
214. Reddy, M.S.S., Chen, F., Shadle, G., Jackson, L., Aljoe, H. & Dixon, R.A. (2005) Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (*Medicago sativa* L.). *Proceedings of the National Academy of Sciences, USA*, **102**, 16573–16578.
215. Jourdes, M., Patten, A.M., Laskar, D.L., Helms, G.L., Franceschi, V.R., Davin, L.B. & Lewis, N.G. (2008) (Manuscript in revision).
216. Dickison, W.C. (2000) *Integrative Plant Anatomy*. Harcourt Academic Press, Burlington, MA.
217. Lapierre, C., Pollet, B., Monties, B. & Rolando, C. (1991) Thioacidolysis of spruce lignin: GC-MS analysis of the main dimers recovered after Raney nickel desulphuration. *Holzforchung*, **45**, 61–68.
218. Terashima, N., Atalla, R.H., Ralph, S.A., Landucci, L.L., Litpierre, C. & Monties, B. (1996) New preparations of lignin polymer models under conditions that approximate cell wall lignification. II. Structural characterization of the models by thioacidolysis. *Holzforchung*, **50**, 9–14.
219. Hoffmann, L., Besseau, S., Geoffroy, P., Ritzenthaler, C., Meyer, D., Lapierre, C., Pollet, B. & Legrand, M. (2004) Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell*, **16**, 1446–1465.
220. Besseau, S., Hoffmann, L., Geoffroy, P., Lapierre, C., Pollet, B. & Legrand, M. (2007) Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell*, **19**, 148–162.

221. Kajita, S., Katayama, Y. & Omori, S. (1996) Alterations in the biosynthesis of lignin in transgenic plants with chimeric genes for 4-coumarate: coenzyme A ligase. *Plant Cell Physiology*, **37**, 957–965.
222. Kajita, S., Hishiyama, S., Tomimura, Y., Katayama, Y. & Omori, S. (1997) Structural characterization of modified lignin in transgenic tobacco plants in which the activity of 4-coumarate: coenzyme A ligase is depressed. *Plant Physiology*, **114**, 871–879.
223. Kajita, S., Mashino, Y., Nishikubo, N., Katayama, Y. & Omori, S. (1997) Immunological characterization of transgenic tobacco plants with a chimeric gene for 4-coumarate:CoA ligase that have altered lignin in their xylem tissue. *Plant Science*, **128**, 109–118.
224. Hu, W.-J., Harding, S.A., Lung, J., Popko, J.L., Ralph, J., Stokke, D.D., Tsai, C.-J. & Chiang, V.L. (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nature Biotechnology*, **17**, 808–812.
225. Pearce, G., Marchand, P.A., Griswold, J., Lewis, N.G. & Ryan, C.A. (1998) Accumulation of feruloyltyramine and *p*-coumaroyltyramine in tomato leaves in response to wounding. *Phytochemistry*, **47**, 659–664.
226. Boudet, A.-M. (1998) A new view of lignification. *Trends in Plant Science*, **3**, 67–71.
227. Porter, K.S., Axtell, J.D., Lechtenberg, V.L. & Colenbrander, V.F. (1978) Phenotype, fiber composition, and *in vitro* dry matter disappearance of chemically induced brown midrib (*bmr*) mutants of sorghum. *Crop Science*, **18**, 205–208.
228. Cherney, J.H., Axtell, J.D., Hassen, M.M. & Anliker, K.S. (1988) Forage quality characterization of a chemically induced brown-midrib mutant in pearl millet. *Crop Science*, **28**, 783–787.
229. Bout, S. & Vermerris, W. (2003) A candidate-gene approach to clone the sorghum *brown midrib* gene encoding caffeic acid *O*-methyltransferase. *Molecular Genetics and Genomics*, **269**, 205–214.
230. Cherney, J.H., Cherney, D.J.R., Akin, D.E. & Axtell, J.D. (1991) Potential of *brown-midrib*, low-lignin mutants for improving forage quality. *Advances in Agronomy*, **46**, 157–198.
231. Grand, C., Parmentier, P., Boudet, A. & Boudet, A.-M. (1985) Comparison of lignins and of enzymes involved in lignification in normal and *brown midrib* (*bm₃*) mutant corn seedlings. *Physiologie Végétale*, **23**, 905–911.
232. Piquemal, J., Lapiere, C., Myton, K., O'Connell, A., Schuch, W., Grima-Pettenati, J. & Boudet, A.-M. (1998) Down-regulation of cinnamoyl-CoA reductase induces significant changes of lignin profiles in transgenic tobacco plants. *Plant Journal*, **13**, 71–83.
233. Jones, L., Ennos, A.R. & Turner, S.R. (2001) Cloning and characterization of *irregular xylem4* (*irx4*): A severely lignin-deficient mutant of *Arabidopsis*. *Plant Journal*, **26**, 205–216.
234. Bernard-Vailhé, M.-A., Cornu, A., Robert, D., Maillot, M.-P. & Besle, J.-M. (1996) Cell wall degradability of transgenic tobacco stems in relation to their chemical extraction and lignin quality. *Journal of Agricultural and Food Chemistry*, **44**, 1164–1169.
235. Halpin, C., Knight, M.E., Foxon, G.A., Campbell, M.M., Boudet, A.M., Boon, J.J., Chabbert, B., Tollier, M.-T. & Schuch, W. (1994) Manipulation of lignin quality by downregulation of cinnamyl alcohol dehydrogenase. *Plant Journal*, **6**, 339–350.
236. Hibino, T., Takabe, K., Kawazu, T., Shibata, D. & Higuchi, T. (1995) Increase of cinnamaldehyde groups in lignin of transgenic tobacco plants carrying an antisense gene for cinnamyl alcohol dehydrogenase. *Bioscience, Biotechnology, and Biochemistry*, **59**, 929–931.
237. Higuchi, T., Ito, T., Umezawa, T., Hibino, T. & Shibata, D. (1994) Red-brown color of lignified tissues of transgenic plants with antisense CAD gene: Wine-red lignin from coniferyl aldehyde. *Journal of Biotechnology*, **37**, 151–158.
238. Baucher, M., Chabbert, B., Pilate, G., Van Doorselaere, J., Tollier, M.-T., Petit-Conil, M., Cornu, D., Monties, B., Van Montagu, M., Inzé, D., Jouanin, L. & Boerjan, W. (1996) Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar. *Plant Physiology*, **112**, 1479–1490.

239. Kim, H., Ralph, J., Lu, F., Ralph, S.A., Boudet, A.-M., MacKay, J.J., Sederoff, R.R., Ito, T., Kawai, S., Ohashi, H. & Higuchi, T. (2003) NMR analysis of lignins in CAD-deficient plants. Part 1. Incorporation of hydroxycinnamaldehydes and hydroxybenzaldehydes into lignins. *Organic and Biomolecular Chemistry*, **1**, 268–281.
240. Kasahara, H., Davin, L.B. & Lewis, N.G. (2004) Aryl propenal double bond reductase. United States Patent No. 6,703,229. Filed March 27, 2001. Issued March 9, 2004, 30 pp.
241. Kasahara, H., Jiao, Y., Bedgar, D.L., Kim, S.-J., Patten, A.M., Xia, Z.-Q., Davin, L.B. & Lewis, N.G. (2006) *Pinus taeda* phenylpropenal double-bond reductase: Purification, cDNA cloning, heterologous expression in *Escherichia coli*, and subcellular localization in *P. taeda*. *Phytochemistry*, **67**, 1765–1780.
242. Youn, B., Kim, S.-J., Moinuddin, S.G.A., Lee, C., Bedgar, D.L., Harper, A.R., Davin, L.B., Lewis, N.G. & Kang, C. (2006) Mechanistic and structural studies of apoform, binary, and ternary complexes of the *Arabidopsis* alkenal double bond reductase At5g16970. *Journal of Biological Chemistry*, **281**, 40076–40088.
243. Yahiaoui, N., Marque, C., Myton, K.E., Negrel, J. & Boudet, A.M. (1998) Impact of different levels of cinnamyl alcohol dehydrogenase down-regulation on lignins of transgenic tobacco plants. *Planta*, **204**, 8–15.
244. Huang, X., Jeronimidis, G. & Vincent, J.F.V. (1999) Mechanical properties of wood from transgenic poplar trees with modified lignification. *The 2nd Symposium of Chinese Youth Scholars on Material Science and Technology*, **10**, 1–9.
245. Tieman, D.M., Loucas, H.M., Kim, J.Y., Clark, D.G. & Klee, H.J. (2007) Tomato phenylacetaldehyde reductases catalyze the last step in the synthesis of the aroma volatile 2-phenylethanol. *Phytochemistry*, **68**, 2660–2669.
246. Chapple, C.C.S., Vogt, T., Ellis, B.E. & Somerville, C.R. (1992) An *Arabidopsis* mutant defective in the general phenylpropanoid pathway. *Plant Cell*, **4**, 1413–1424.
247. Meyer, K., Shirley, A.M., Cusumano, J.C., Bell-Lelong, D.A. & Chapple, C. (1998) Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA*, **95**, 6619–6623.
248. Marita, J.M., Ralph, J., Hatfield, R.D. & Chapple, C. (1999) NMR characterization of lignins in *Arabidopsis* altered in the activity of ferulate 5-hydroxylase. *Proceedings of the National Academy of Sciences, USA*, **96**, 12328–12332.
249. Ralph, J., Lapierre, C., Marita, J.M., Kim, H., Lu, F., Hatfield, R.D., Ralph, S., Chapple, C., Franke, R., Hemm, M.R., Van Doorselaere, J., Sederoff, R.R., O'Malley, D.M., Scott, J.T., MacKay, J.J., Yahiaoui, N., Boudet, A.-M., Pean, M., Pilate, G., Jouanin, L. & Boerjan, W. (2001) Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. *Phytochemistry*, **57**, 993–1003.
250. Zuber, M.S., Colbert, T.R. & Bauman, L.F. (1977) Effect of *brown-midrib-3* mutant in maize (*Zea mays* L.) on stalk strength. *Zeitschrift für Pflanzenzüchtung*, **79**, 310–314.
251. Tsai, C.-J., Popko, J.L., Mielke, M.R., Hu, W.-J., Podila, G.K. & Chiang, V.L. (1998) Suppression of *O*-methyltransferase gene by homologous sense transgene in quaking aspen causes red-brown wood phenotypes. *Plant Physiology*, **117**, 101–112.
252. Guo, D., Chen, F., Inoue, K., Blount, J.W. & Dixon, R.A. (2001) Downregulation of caffeic acid 3-*O*-methyltransferase and caffeoyl CoA 3-*O*-methyltransferase in transgenic alfalfa: Impacts on lignin structure and implications for the biosynthesis of G and S lignin. *Plant Cell*, **13**, 73–88.
253. Marita, J.M., Ralph, J., Hatfield, R.D., Guo, D., Chen, F. & Dixon, R.A. (2003) Structural and compositional modifications in lignin of transgenic alfalfa down-regulated in caffeic acid 3-*O*-methyltransferase and caffeoyl coenzyme A 3-*O*-methyltransferase. *Phytochemistry*, **62**, 53–65.
254. Higuchi, T. (1957) Biochemical studies of lignin formation. I. *Physiologia Plantarum*, **10**, 356–372.

255. Harkin, J.M. & Obst, J.R. (1973) Lignification in trees: Indication of exclusive peroxidase participation. *Science*, **180**, 296–297.
256. Ros Barceló, A. (1995) Peroxidase and not laccase is the enzyme responsible for cell wall lignification in the secondary thickening of xylem vessels in *Lupinus*. *Protoplasma*, **186**, 41–44.
257. Blee, K.A., Choi, J.W., O'Connell, A.P., Schuch, W., Lewis, N.G. & Bolwell, G.P. (2003) A lignin-specific peroxidase in tobacco whose antisense suppression leads to vascular tissue modification. *Phytochemistry*, **64**, 163–176.
258. Manskaya, S.M. (1948) Participation of oxidases in lignin formation. *Doklady Akademii Nauk SSSR*, **62**, 369–371.
259. Manskaya, S.M. (1948) The formation of lignin in plants. *Trudy Konferentsiya Vysokomolekulyarnym Soedineniyam*, 158–171.
260. Higuchi, T. (1958) Further studies on phenol oxidase related to the lignin biosynthesis. *Journal of Biochemistry*, **45**, 515–528.
261. Bligny, R. & Douce, R. (1983) Excretion of laccase by sycamore (*Acer pseudoplatanus* L.) cells. *Biochemical Journal*, **209**, 489–496.
262. Driouich, A., Lainé, A.-C., Vian, B. & Faye, L. (1992) Characteration and localization of laccase forms in stem and cell cultures of sycamore. *Plant Journal*, **2**, 13–24.
263. Sterjiades, R., Dean, J.F.D. & Eriksson, K.-E.L. (1992) Laccase from sycamore maple (*Acer pseudoplatanus*) polymerizes monolignols. *Plant Physiology*, **99**, 1162–1168.
264. Bao, W., O'Malley, D.M., Whetten, R. & Sederoff, R.R. (1993) A laccase associated with lignification in loblolly pine xylem. *Science*, **260**, 672–674.
265. LaFayette, P.R., Eriksson, K.-E.L. & Dean, J.F.D. (1995) Nucleotide sequence of a cDNA clone encoding an acidic laccase from sycamore maple (*Acer pseudoplatanus* L.). *Plant Physiology*, **107**, 667–668.
266. Liang, M., Davis, E., Gardner, D., Cai, X. & Wu, Y. (2006) Involvement of *AtLAC15* in lignin synthesis in seeds and in root elongation of *Arabidopsis*. *Planta*, **224**, 1185–1196.
267. Sterjiades, R., Dean, J.F.D., Gamble, G., Himmelsbach, D.S. & Eriksson, K.-E.L. (1993) Extracellular laccases and peroxidases from sycamore maple (*Acer pseudoplatanus*) cell-suspension cultures. *Planta*, **190**, 75–87.
268. Savidge, R. & Udagama-Randeniya, P. (1992) Cell-wall bound coniferyl alcohol oxidase associated with lignification in conifers. *Phytochemistry*, **31**, 2959–2966.
269. Udagama-Randeniya, P. & Savidge, R. (1994) Electrophoretic analysis of coniferyl alcohol oxidase and related laccases. *Electrophoresis*, **15**, 1072–1077.
270. McDougall, G.J., Stewart, D. & Morrison, I.M. (1994) Cell-wall-bound oxidases from tobacco (*Nicotiana tabacum*) xylem participate in lignin formation. *Planta*, **194**, 9–14.
271. Richardson, A., Stewart, D. & McDougall, G.J. (1997) Identification and partial characterization of a coniferyl alcohol oxidase from lignifying xylem of Sitka spruce (*Picea sitchensis*). *Planta*, **203**, 35–43.
272. Freudenberg, K. & Richtzenhain, H. (1943) Enzymatische Versuche zur Entstehung des Lignins. *Chemische Berichte*, **76**, 997–1006.
273. Freudenberg, K. & Heimberger, W. (1950) Die biochemische Synthese ligninartiger Stoffe. *Chemische Berichte*, **83**, 519–530.
274. Freudenberg, K., Kraft, R. & Heimberger, W. (1951) Über den Sinapinalkohol, den Coniferylalkohol und ihre Dehydrierungspolymerisate. *Chemische Berichte*, **84**, 472–476.
275. Freudenberg, K., Reznik, H., Boesenberg, H. & Rasenack, D. (1952) Das an der Verholzung beteiligte Fermentsystem. *Chemische Berichte*, **85**, 641–647.
276. Freudenberg, K. & Dietrich, H. (1953) Synthese des *d,l*-Pinoresinols und andere Versuche im Zusammenhang mit dem Lignin. *Chemische Berichte*, **86**, 1157–1167.
277. Freudenberg, K. & Bittner, F. (1953) Versuche mit Coniferylalkohol, der radioaktiven Kohlenstoff enthält. *Chemische Berichte*, **86**, 155–159.

278. Mason, H.S. & Cronyn, M. (1955) On the role of polyphenoloxidase in lignin biosynthesis. *Journal of the American Chemical Society*, **77**, 491–492.
279. Koblit, H. & Koblit, D. (1964) Participation of cytochrome oxidase in lignification. *Nature*, **204**, 199–200.
280. Ros Barceló, A. (1997) Lignification in plant cell walls. *International Review of Cytology*, **176**, 87–132.
281. Ranocha, P., Chabannes, M., Chamayou, S., Danoun, S., Jauneau, A., Boudet, A.-M. & Goffner, D. (2002) Laccase down-regulation causes alterations in phenolic metabolism and cell wall structure in poplar. *Plant Physiology*, **129**, 145–155.
282. Pourcel, L., Routaboul, J.-M., Kerhoas, L., Caboche, M., Lepiniec, L. & Debeaujon, I. (2005) *TRANSPARENT TESTA10* encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat. *Plant Cell*, **17**, 2966–2980.
283. Hiratsu, K., Mitsuda, N., Matsui, K. & Ohme-Takagi, M. (2004) Identification of the minimal repression domain of *SUPERMAN* shows that the DLELRL hexapeptide is both necessary and sufficient for repression of transcription in *Arabidopsis*. *Biochemical and Biophysical Research Communications*, **321**, 172–178.
284. Burlat, V., Kwon, M., Davin, L.B. & Lewis, N.G. (2001) Dirigent proteins and dirigent sites in lignifying tissues. *Phytochemistry*, **57**, 883–897.
285. Davin, L.B. & Lewis, N.G. (2000) Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiology*, **123**, 453–461.
286. Lewis, N.G. (1999) A 20th century roller coaster ride: A short account of lignification. *Current Opinion in Plant Biology*, **2**, 153–162.
287. Freudenberg, K. & Schlüter, H. (1955) Weitere Zwischenprodukte der Ligninbildung. *Chemische Berichte*, **88**, 617–625.
288. Freudenberg, K. & Hübner, H.H. (1952) Oxyzimtalkohole und ihre Dehydrierungs-polymerisate. *Chemische Berichte*, **85**, 1181–1191.
289. Freudenberg, K., Harkin, J.M., Reichert, M. & Fukuzumi, T. (1958) Die an der Verholzung beteiligten Enzyme. Die Dehydrierung des Sinapinalkohols. *Chemische Berichte*, **91**, 581–590.
290. Freudenberg, K. (1966) Analytical and biochemical background of a constitutional scheme of lignin. *Advances in Chemistry Series*, **59**, 1–21.
291. Freudenberg, K. & Harkin, J.M. (1964) Ergänzung des Konstitutionsschemas für das Lignin der Fichte. *Holzforschung*, **18**, 166–168.
292. Freudenberg, K. (1964) Entwurf eines Konstitutionsschemas für das Lignin der Fichte. *Holz-forschung*, **18**, 3–9.
293. Dovgan, I.V. & Medvedeva, E.I. (1983) Change in the structural elements of the lignin of the brown alga *Cystoseira barbata* at different ages. *Chemistry of Natural Compounds*, **19**, 81–84.
294. Dovgan, I.V., Medvedeva, E.I. & Yanishevskaya, E.N. (1983) Cleavage of the lignins of the alga *Cystoseira barbata* by thioacetic acid. *Chemistry of Natural Compounds*, **19**, 84–87.
295. Reznikov, V.M., Mikhaseva, M.F. & Zil'bergleit, M.A. (1978) The lignin of the alga *Fucus vesiculosus*. *Chemistry of Natural Compounds*, **14**, 554–556.
296. Foster, J.W. (1949) *Chemical Activities of Fungi*. Academic Press, New York.
297. Freudenberg, K. (1955) Lignin. In: *Moderne Methoden der Pflanzenanalyse* (eds. K. Paech & M.V. Tracey), pp. 499–516. Springer-Verlag, Berlin, Germany.
298. Bu'Lock, J.D. & Smith, H.G. (1961) A fungus pigment of novel type, and the nature of fungus "lignin". *Experientia*, **XVII**, 553–554.
299. Bu'Lock, J.D., Leeming, P.R. & Smith, H.G. (1962) Pyrones. Part II. Hispidin, a new pigment and precursor of a fungus "lignin". *Journal of the Chemical Society*, 2085–2089.
300. Tutschek, R. (1975) Isolation and characterization of the *p*-hydroxy- β -(carboxymethyl)-cinnamic acid (sphagnum acid) from the cell wall of *Sphagnum magellanicum* Brid. *Zeitschrift für Pflanzenphysiologie*, **76**, 353–365.

301. Ragan, M.A. (1984) *Fucus* "lignin": A reassessment. *Phytochemistry*, **23**, 2029–2032.
302. Lewis, N.G. & Davin, L.B. (1998) The biochemical control of monolignol coupling and structure during lignan and lignin biosynthesis. In: *Lignin and Lignan Biosynthesis*, Vol. 697 (eds. N.G. Lewis & S. Sarkanen), pp. 334–361. ACS Symposium Series, Washington, DC.
303. Becker, H. & Nimz, H. (1974) Investigations of lignin from European mistletoe (*Viscum album* L.) in dependence from its host. *Zeitschrift für Pflanzenphysiologie*, **72**, 52–63.
304. Goldschmid, O. & Hergert, H.L. (1961) Examination of western hemlock for lignin precursors. *Tappi*, **44**, 858–870.
305. Marcinowski, S. & Grisebach, H. (1977) Turnover of coniferin in pine seedlings. *Phytochemistry*, **16**, 1665–1667.
306. Forss, K. & Fremer, K.-E. (1964) The dissolution of wood components under different conditions of sulfite pulping. *Tappi*, **47**, 485–493.
307. Forss, K. & Fremer, K.-E. (1965) The repeating unit in spruce lignin. *Paperi ja Puu*, **47**, 443–454.
308. Forss, K. & Fremer, K.-E. (1983) Comments on the nature of coniferous lignin. *Journal of Applied Polymer Science: Applied Polymer Symposium*, **37**, 531–547.
309. Nimz, H. (1974) Beech lignin—proposal of a constitutional scheme. *Angewandte Chemie (International ed. in English)*, **13**, 313–321.
310. Nimz, H. (1969) Degradation of lignin. *Chemische Berichte*, **102**, 799–810.
311. Önnnerud, H., Palmblad, M. & Gellerstedt, G. (2003) Investigation of lignin oligomers using electrospray ionisation mass spectrometry. *Holzforschung*, **57**, 37–43.
312. Adler, E. (1977) Lignin chemistry. Past, present and future. *Wood Science and Technology*, **11**, 169–218.
313. Sakakibara, A. (1980) A structural model of softwood lignin. *Wood Science and Technology*, **14**, 89–100.
314. Banoub, J.H. & Delmas, M. (2003) Structural elucidation of the wheat straw lignin polymer by atmospheric pressure chemical ionization tandem mass spectrometry and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, **38**, 900–903.
315. Lam, H.Q., Le Bigot, Y., Delmas, M. & Avignon, G. (2001) A new procedure for the destructuring of vegetable matter at atmospheric pressure by a catalyst/solvent system of formic acid/acetic acid applied to the pulping of triticale straw. *Industrial Crops and Products*, **14**, 139–144.
316. Dolk, M., Pla, F., Yan, J.F. & McCarthy, J.L. (1986) Lignin. 22. Macromolecular characteristics of alkali lignin from western hemlock wood. *Macromolecules*, **19**, 1464–1470.
317. Mlynár, J. & Sarkanen, S. (1996) Renaissance in ultracentrifugal sedimentation equilibrium calibrations of size exclusion chromatographic elution profiles. In: *Strategies in Size Exclusion Chromatography*, Vol. 635 (eds. M. Potschka & P.L. Dubin), pp. 379–400. ACS Symposium Series, Washington, DC.
318. Smith, D.C., Glasser, W.G., Glasser, H.R. & Ward, T.C. (1988) Simulation of reactions with lignin by computer (SIMREL). VII. About the gel structure of native lignin. *Cellulose Chemistry and Technology*, **22**, 171–190.
319. Hatakeyama, T., Yoshida, T. & Hatakeyama, H. (1999) Thermal and viscoelastic properties of *in situ* lignin. *10th International Symposium on Wood and Pulp Chemistry*, **1**, 478–481.
320. Li, Y. & Sarkanen, S. (2003) Biodegradable kraft lignin-based thermoplastics. In: *Biodegradable Polymers and Plastics* (eds. E. Chiellini & R. Solaro). Kluwer Academic/Plenum Publishers, New York.
321. Lewis, N.G. & Davin, L.B. (1999) Lignans: Biosynthesis and function. In: *Comprehensive Natural Products Chemistry*, Vol. 1 (eds. Sir D.H.R. Barton, K. Nakanishi & O. Meth-Cohn), pp. 639–712. Elsevier, Oxford.
322. Davin, L.B. & Lewis, N.G. (2003) An historical perspective on lignan biosynthesis: Monolignol, allylphenol and hydroxycinnamic acid coupling and downstream metabolism. *Phytochemistry Reviews*, **2**, 257–288.

323. Davin, L.B., Bedgar, D.L., Katayama, T. & Lewis, N.G. (1992) On the stereoselective synthesis of (+)-pinoreosinol in *Forsythia suspensa* from its achiral precursor, coniferyl alcohol. *Phytochemistry*, **31**, 3869–3874.
324. Davin, L.B., Wang, H.-B., Crowell, A.L., Bedgar, D.L., Martin, D.M., Sarkanen, S. & Lewis, N.G. (1997) Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. *Science*, **275**, 362–366.
325. Gang, D.R., Costa, M.A., Fujita, M., Dinkova-Kostova, A.T., Wang, H.-B., Burlat, V., Martin, W., Sarkanen, S., Davin, L.B. & Lewis, N.G. (1999) Regiochemical control of monolignol radical coupling: A new paradigm for lignin and lignan biosynthesis. *Chemistry and Biology*, **6**, 143–151.
326. Halls, S.C. & Lewis, N.G. (2002) Secondary and quaternary structures of the (+)-pinoreosinol-forming dirigent protein. *Biochemistry*, **41**, 9455–9461.
327. Halls, S.C., Davin, L.B., Kramer, D.M. & Lewis, N.G. (2004) Kinetic study of coniferyl alcohol radical binding to the (+)-pinoreosinol forming dirigent protein. *Biochemistry*, **43**, 2587–2595.
328. Cho, M.-H., Moinuddin, S.G.A., Helms, G.L., Hishiyama, S., Eichinger, D., Davin, L.B. & Lewis, N.G. (2003) (+)-Larreatricin hydroxylase, an enantio-specific polyphenol oxidase from the creosote bush (*Larrea tridentata*). *Proceedings of the National Academy of Sciences, USA*, **100**, 10641–10646.
329. Moinuddin, S.G.A., Hishiyama, S., Cho, M.-H., Davin, L.B. & Lewis, N.G. (2003) Synthesis and chiral HPLC analysis of the dibenzyltetrahydrofuran lignans, larreatricins, 8'-*epi*-larreatricins, 3,3'-didemethoxyverrucosins and *meso*-3,3'-didemethoxynectandrin B in the creosote bush (*Larrea tridentata*): Evidence for regiospecific control of coupling. *Organic and Biomolecular Chemistry*, **1**, 2307–2313.
330. Kim, M.K., Jeon, J.-H., Davin, L.B. & Lewis, N.G. (2002) Monolignol radical–radical coupling networks in western red cedar and *Arabidopsis* and their evolutionary implications. *Phytochemistry*, **61**, 311–322.
331. Ralph, S., Park, J.-Y., Bohlmann, J. & Mansfield, S.D. (2006) Dirigent proteins in conifer defense: Gene discovery, phylogeny, and differential wound- and insect-induced expression of a family of DIR and DIR-like genes in spruce (*Picea* spp.). *Plant Molecular Biology*, **60**, 21–40.
332. Ralph, S.G., Jancsik, S. & Bohlmann, J. (2007) Dirigent proteins in conifer defense. II. Extended gene discovery, phylogeny, and constitutive and stress-induced gene expression in spruce (*Picea* spp.). *Phytochemistry*, **68**, 1975–1991.
333. Kim, M.K., Jeon, J.-H., Fujita, M., Davin, L.B. & Lewis, N.G. (2002) The western red cedar (*Thuja plicata*) 8–8' DIRIGENT family displays diverse expression patterns and conserved monolignol coupling specificity. *Plant Molecular Biology*, **49**, 199–214.
334. Kim, K.-W., Franceschi, V.R., Davin, L.B. & Lewis, N.G. (2006) β -Glucuronidase as reporter gene: Advantages and limitations. In: *Methods in Molecular Biology. Vol. 323: Arabidopsis Protocols*, 2nd edn (eds J. Salinas & J.J. Sanchez-Serrano), pp. 263–273. Humana Press, Totowa.
335. Lourith, N., Katayama, T., Ishikawa, K. & Suzuki, T. (2005) Biosynthesis of a syringyl 8-*O*-4' neolignan in *Eucommia ulmoides*: Formation of syringylglycerol-8-*O*-4'-(sinapyl alcohol) ether from sinapyl alcohol. *Journal of Wood Science*, **51**, 379–386.
336. Lourith, N., Katayama, T. & Suzuki, T. (2005) Stereochemistry and biosynthesis of 8-*O*-4' neolignans in *Eucommia ulmoides*: Diastereoselective formation of guaiacylglycerol-8-*O*-4'-(sinapyl alcohol) ether. *Journal of Wood Science*, **51**, 370–378.
337. Liu, J., Bell, A.A., Stipanovic, R.D. & Puckhaber, L.S. (2008) Increasing cottonseed utilization through breeding and genetic engineering to produce high levels of (+)-gossypol in seed. (submitted).
338. Niemetz, R. & Gross, G.G. (2003) Ellagitannin biosynthesis: Laccase-catalyzed dimerization of tellimagrandin II to cornusiin E in *Tellima grandiflora*. *Phytochemistry*, **64**, 1197–1201.
339. Niemetz, R. & Gross, G.G. (2003) Oxidation of pentagalloylglucose to the ellagitannin, tellimagrandin II, by a phenol oxidase from *Tellima grandiflora* leaves. *Phytochemistry*, **62**, 301–306.

340. Niemetz, R. & Gross, G.G. (2005) Enzymology of gallotannin and ellagitannin biosynthesis. *Phytochemistry*, **66**, 2001–2011.
341. Ralph, J., Peng, J., Lu, F., Hatfield, R.D. & Helm, R.F. (1999) Are lignins optically active? *Journal of Agricultural and Food Chemistry*, **47**, 2991–2996.
342. Cho, M.H., Milhollan, J.K., Lane, W.S., Lange, B.M., Davin, L.B. & Lewis, N.G. (2008) The proteomes of the vascular apparatus and associated cell types in *Arabidopsis thaliana*. (Manuscript submitted).
343. Müsel, G., Schindler, T., Bergfeld, R., Ruel, K., Jacquet, G., Lapierre, C., Speth, V. & Schopfer, P. (1997) Structure and distribution of lignin in primary and secondary cell walls of maize coleoptiles analyzed by chemical and immunological probes. *Planta*, **201**, 146–159.
344. Ryser, U., Schorderet, M., Zhao, G.-F., Studer, D., Ruel, K., Hauf, G. & Keller, B. (1997) Structural cell-wall proteins in protoxylem development: Evidence for a repair process mediated by a glycine-rich protein. *Plant Journal*, **12**, 97–111.
345. Zhang, L., Henriksson, G. & Gellerstedt, G. (2003) The formation of β - β structures in lignin biosynthesis—are there two different pathways? *Organic and Biomolecular Chemistry*, **1**, 3621–3624.
346. Dinkova-Kostova, A.T., Gang, D.R., Davin, L.B., Bedgar, D.L., Chu, A. & Lewis, N.G. (1996) (+)-Pinoresinol/(+)-lariciresinol reductase from *Forsythia intermedia*: Protein purification, cDNA cloning, heterologous expression and comparison to isoflavone reductase. *Journal of Biological Chemistry*, **271**, 29473–29482.
347. Fujita, M., Gang, D.R., Davin, L.B. & Lewis, N.G. (1999) Recombinant pinoresinol-lariciresinol reductases from western red cedar (*Thuja plicata*) catalyze opposite enantiospecific conversions. *Journal of Biological Chemistry*, **274**, 618–627.
348. Min, T., Kasahara, H., Bedgar, D.L., Youn, B., Lawrence, P.K., Gang, D.R., Halls, S.C., Park, H., Hilsenbeck, J.L., Davin, L.B., Lewis, N.G. & Kang, C. (2003) Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. *Journal of Biological Chemistry*, **278**, 50714–50723.
349. Ralph, J. & Brunow, G. (2006) Lignification: Are lignins biosynthesized via simple combinatorial chemistry or via proteinaceous control and template replication? *Communication: XXIII International Conference on Polyphenols*, Winnipeg, Manitoba, Canada, August 22–25.
350. Holmgren, A., Brunow, G., Henriksson, G., Zhang, L. & Ralph, J. (2006) Non-enzymatic reduction of quinone methides during oxidative coupling of monolignols: Implications for the origin of benzyl structures in lignins. *Organic and Biomolecular Chemistry*, **4**, 3456–3461.
351. Samuels, A.L., Rensing, K.H., Douglas, C.J., Mansfield, S.D., Dharmawardhana, D.P. & Ellis, B.E. (2002) Cellular machinery of wood production: Differentiation of secondary xylem in *Pinus contorta* var. *latifolia*. *Planta*, **216**, 72–82.
352. Önnerud, H., Zhang, L., Gellerstedt, G. & Henriksson, G. (2002) Polymerization of monolignols by redox shuttle-mediated enzymatic oxidation: A new model in lignin biosynthesis. *Plant Cell*, **14**, 1953–1962.
353. Chen, Y.-r. & Sarkanen, S. (2007) Noncovalent interactions between lignin substructures – implications for lignin biosynthesis and effects on properties of lignin-based polymeric materials. *Proceedings of the 8th International Lignin Institute Forum*, Rome, May 10–12, pp. 1–4.

Chapter 8

Computational Approaches to Study Cellulose Hydrolysis

Michael F. Crowley and Ross C. Walker

8.1 Introduction

Molecular modeling is a process employing a powerful set of tools for probing the atomistic mechanisms of cellulose hydrolysis, and will feature prominently in efforts to harness cellulose as a biomass energy resource. It provides a bridge between theoretical concepts and proposed mechanisms and the experimental data. Molecular modeling is most powerful when used in a synergistic fashion with experimentation where hypothesis-driven computational research is used to help explain theoretical observations and to drive the design of future experiments. Molecular modeling encompasses the entire range of computational approaches available to the molecular biologist and it is beyond the scope of this book to give a comprehensive overview of all the computational models that exist. Instead we will concentrate on the subset of molecular modeling termed molecular dynamics, which will prove crucial in advancing our understanding of the behavior of celluloses and cellulases on the atomistic scale.

Molecular dynamics (MD) is generally used as a virtual experimental tool to probe the structure, function, kinetic, and thermodynamic properties of substances. In the biomolecular field, it has been invaluable in validating structures and elucidating mechanisms of structural stability and conformational change and for understanding interactions between molecules, their ligands, and their constituent parts. Most of MD is based on classical molecular mechanics (MM) with a smaller amount of work on using quantum mechanics (QM) with molecular mechanics to produce hybrid, QM/MM, dynamics methods. Although many of the computational methods used in molecular dynamics studies of biomolecular systems are mature, having been extensively applied to proteins, small molecules, and to nucleic acids, there has been, until recently comparatively little interest in the use of MD methods for carbohydrates and even less so for cellulose. The thrust of this chapter will be to document the current state of MD methods and MM force fields with the intent of inspiring greater use of the tools for the study of cellulosic recalcitrance to complement experimental studies, to answer questions that are unapproachable by current experimental technology, and to provide to experimentalists a wish list of new experimental targets, mutations, and structural information. We will include the work already accomplished, and outline the currently available methods and the kinds of questions they can answer for systems of the

size, complexity, and chemical nature of cellulose and the enzymes and other biomolecules that interact with it.

8.2 Molecular mechanics

Ab initio quantum mechanical potentials offer more rigorous descriptions of molecules, than classical potentials, but, despite recent advances, require computational resources that are far greater than what is currently available for systems, the size of proteins and celluloses. Thus for the study of such systems at an atomic level, force field methods present the only practical technique widely available. In contrast to quantum mechanics calculations, force field (or molecular mechanics) methods do not take into account electronic effects, considering only nuclear motions. Such a separation is facilitated by the Born–Oppenheimer approximation (1) which asserts that, due to the significant difference in mass, the motion of electrons is decoupled from the motion of nuclei. The separation allows the nuclear positions to be described by classical Newtonian mechanics. The use of classical mechanics greatly reduces the complexity of the calculations involved in simulating large molecular systems and makes possible simulations involving hundreds of thousands of atoms. The drawback of the force field approach is that electronic effects cannot be examined. Calculations are therefore restricted to the ground state, although excited state MD simulations have been attempted using specially parameterized force fields (2). A classical approach also excludes the direct study of chemical reactions or other phenomena where changes in electron distribution are significant. In order to study large systems in which electronic effects are relevant it is necessary to use combined quantum mechanical/molecular mechanical (QM/MM) approaches as discussed later in this chapter.

8.2.1 The force field equation

It is difficult to accept that the behavior of atomistic systems, which behave according to quantum rather than classical laws, could be accurately described by the application of classical Newtonian mechanics. This approach can be justified, however, by considering the de Broglie expression for the thermal wavelength- Λ ,

$$\Lambda = \sqrt{\frac{2\pi\hbar^2}{Mk_B T}} \quad (8.1)$$

where T is the temperature and M is the atomic mass. The approximation of classical behavior holds if $\Lambda = \alpha$, where α is the mean nearest neighbor separation. This holds for “heavy” liquid systems at all but the lowest temperatures at which quantum effects become important.

To describe the molecular dynamics of a system classically a function representing the potential energy of the system, together with the related parameters is required. Typically, the energy is calculated from the sum of bonded and non-bonded interactions,

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{non-bond}} \quad (8.2)$$

The exact form of the terms in the above potential function and the associated parameters varies across different molecular mechanical force fields. Some force fields also include a cross term representing coupling between the first three terms in Equation (8.2). Examples of commonly used force fields include the Allinger MM2 and MM3 series (3, 4), CHARMM (5), AMBER (6), and GROMOS (7). Each force field has a slightly different ethos and is typically suited to the study of one class of molecules. Of those mentioned above CHARMM, AMBER, and GROMOS would be considered protein force fields suited to the simulation of proteins, nucleic acids, and with suitable parameterization, carbohydrates. The underlying equations for these three force fields are similar, although there are subtle differences. For the purposes of discussion here we will restrict ourselves to the AMBER force field equation, which is

$$V(r^n) = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon_r R_{ij}} \right] \quad (8.3)$$

where the potential energy V is a function of the positions r of n atoms. K_r , r_{eq} , K_θ , θ_{eq} , V_n , n , ϕ , γ , A_{ij} , B_{ij} , ϵ_r , q_i , and q_j are all empirically defined parameters. The first three terms in the above equation correspond to the bond, angle, and dihedral terms, respectively, while the last term describes the non-bonded van der Waals and electrostatic interactions.

8.2.2 Interatomic potentials

8.2.2.1 The stretch energy

In the AMBER force field there are three terms describing the interactions between atoms that are either directly bonded or separated by two or three bonds. These three terms are known as the bonded terms and correspond to bonds, angles, and dihedrals, respectively. The first term ($K_r (r - r_{eq})^2$) describes a simple harmonic potential obeying Hooke's law and is used to represent the energy involved in the stretching of a bond between two directly linked atoms. $2K_r$ is the force constant for the bond and $(r - r_{eq})$ the distortion from the equilibrium bond length. It should be immediately obvious from this expression that such an approach prohibits the study of bond breaking since the expression tends to infinity as the bond length is increased significantly beyond the equilibrium value.

The harmonic representation of the potential is the simplest possible but provides a fair description of the energetics of bond stretching and compression when the bond length remains close to the equilibrium value (Figure 8.1). While more complex descriptions are available they come at the price of increased computational complexity and increased parameterization requirements and are rarely used for modeling biological systems.

8.2.2.2 The bend energy

The second term in Equation (8.3) describes the angle bend energy between three covalently bound atoms. This also utilizes Hooke's law using the same harmonic potential used for

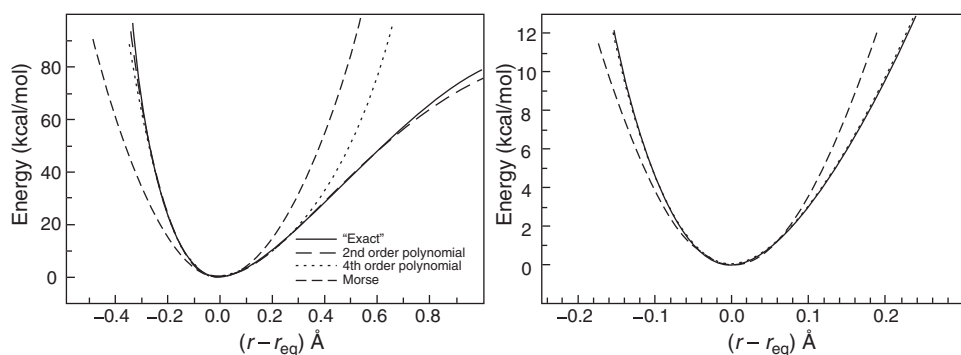


Figure 8.1 The stretch energy for CH_4 showing the various functional forms in comparison to a CASSCF/6-311++G(2df,2pd) electronic structure calculation ("exact"). [Adapted from Jensen (13).]

bond stretching. Here θ_{eq} is the equilibrium bond angle and K_θ the angular force constant. The angular force constants are typically smaller in magnitude than the stretching force constants because the energy penalty for angle bending is generally lower. As with the bond stretching term, more complex descriptions are possible but rarely used.

8.2.2.3 The torsional energy

The bond-stretching and angle-bending terms, discussed above, are often referred to as "hard" degrees of freedom since large energies are required to cause significant deviations from the equilibrium geometries. Most of the complex variations in structure and relative energies observed in biological systems are due to the "softer" torsional and non-bonded contributions.

The barriers to rotation about a bond can be modeled in one of two ways. In very early force fields it was believed that rotational barriers could be omitted. The gauche-trans energy differences would be reproduced by the non-bonded interactions. However, it was quickly realized that for organic molecules neglecting dihedrals made successful parameterization of force fields, to reproduce experimental observables, an almost impossible task and so dihedral terms were explicitly included. The AMBER force field, in common with a number of other biologically oriented force fields, uses a Fourier series expansion for the torsional potential

$$V(\phi) = \sum_{n=0}^N \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \quad (8.4)$$

where V_n is the relative barrier height to rotation, n is the multiplicity (number of minima in a 360° rotation), ϕ is the dihedral angle, and γ is the phase factor which determines the location of the minima. V_n is often termed the relative barrier height since other terms in the force field equation contribute to the barrier height as the bond is rotated, especially the 1-4 non-bonded interactions discussed below. The advantage of using a Fourier series expansion for the dihedral terms centers on the fact that terms of differing multiplicity can be combined to describe complex torsional profiles (Figure 8.2).

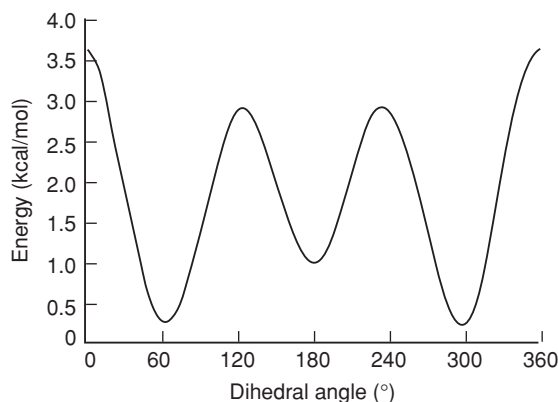


Figure 8.2 Variation in torsional energy with O–C–C–O torsion angle for an OCH₂–CH₂O fragment. [Adapted from Leach (12).]

Improper torsion angles, also known as out-of-plane bending, are defined for four atoms that are not bonded in a serial manner. They are used to maintain planarity where necessary. The AMBER force field accounts for improper torsions in the same way as regular torsion angles but using a twofold multiplicity.

8.2.2.4 Non-bonded interactions

The accurate treatment of non-bonded interactions is paramount in a force field description. The philosophy behind the AMBER force field is to deal with these interactions in the most accurate way possible within the limitations of the available hardware. Non-bonded interactions do not depend upon a specific bonding relationship between atoms. They are “through-space” interactions, the number of which scales roughly as the square of the number of atoms. Unsurprisingly, the non-bonded interactions form the most time-consuming component of molecular mechanics simulations. Molecular mechanics force fields typically consider the non-bonded interactions as two groups, one comprising electrostatic interactions and the other van der Waals interactions.

8.2.2.4.1 ELECTROSTATIC INTERACTIONS

The last term in Equation (8.3) describes the electrostatic interactions within the system. There are a number of ways to represent the charge distribution within a molecule, the simplest being the use of point charges. This is the method utilized by most biologically oriented force fields. In the point charge model a series of fractional charges are distributed throughout the molecule. If the charges are centered on atoms then they are referred to as partial atomic charges. The interaction energy is calculated using Coulomb's law.

Since the charge on an atom is not experimentally observable the partial atomic charges have to be assigned in an analogous fashion to the parameters used in the bonding interaction terms. The partial charges are generally obtained by fitting to a charge potential obtained from an *ab initio* electronic structure calculation.

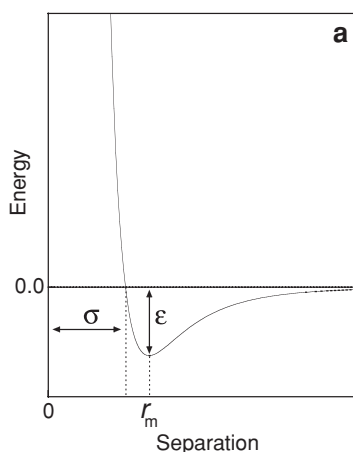


Figure 8.3 The Lennard–Jones 12–6 potential and associated parameters ϵ , r_m , and σ .

8.2.2.4.2 VAN DER WAALS INTERACTIONS

Electrostatics alone cannot account for all non-bonded interactions. Good examples are provided by the noble gases whose multipole moments are all zero, and yet have liquid and solid phases. The forces responsible for this are termed van der Waals forces. These forces are electrostatic in nature, but the magnitude of the interactions is much smaller. Figure 8.3 shows a typical potential energy curve due to van der Waals interactions. This curve arises because of a combination of attractive and repulsive forces. The repulsive term originates from the Pauli exclusion principle such that at very short distances the energy rises steeply but at larger separations the decay is exponential in nature. The attractive term, which dominates at larger distances, is a quantum phenomenon termed dispersion.

A number of expressions exist for the modeling of van der Waals forces. Arguably the best known of these, and the one employed by most biologically oriented force fields is the Lennard–Jones 12–6 function,

$$V(r) = \frac{A}{r^{12}} - \frac{B}{r^6} \quad (8.5)$$

where $A = \epsilon r_m^{12}$ and $B = \epsilon r_m^6$. The Lennard–Jones potential thus contains two adjustable parameters, the well depth ϵ and the minimum energy location r_m (Figure 8.3). It should be noted that while the r^6 term has a clear physical basis the r^{12} term is completely arbitrary and chosen for ease of computation.

8.2.3 Non-bonded cutoffs and long range electrostatics

It is the non-bonded interaction terms that are the most computationally demanding aspect of a force field calculation. There are $N(N-1)/2$ interactions, where N is the number of atoms in the system. Bulk solution is represented most commonly by using a periodic boundary representation in which the unit cell is replicated infinitely in three dimensions. In

this case, the number of interactions for atoms in the primary cell becomes infinite and the standard pairwise electrostatic interaction term becomes a divergent sum. A reduction in the number of non-bonded interactions is thus required in order to make the computation tractable. Since the size of the van der Waals interaction between atoms decreases rapidly with distance it is possible to truncate the Lennard–Jones potential without introducing significant errors in the calculation. Unfortunately, the electrostatic interactions are longer ranged and truncating them can introduce significant errors into the calculation. Much effort has been expended over the years to develop effective cutoff methods that allow the electrostatic interaction to be truncated at some distance, typically below 15 Å. However, all these methods suffer from problems arising due to cancellation of errors and it is now almost universally accepted that cutoffs should not be used unless a method is used which allows the “missing” energy to be calculated. One such method which is now commonly used in explicit solvent simulations is the Particle Mesh Ewald Method (PME) (8), which divides the electrostatic calculation into a direct space, pairwise evaluation, and a reciprocal space calculation. The direct space part of the calculation is conducted using a regular pairwise interaction within a cutoff, typically 8–10 Å while the remainder of the “missing” electrostatic contribution from the infinitely replicated system is included by calculating the charge field on a grid and then using Fast Fourier Transforms to obtain the potential and force at each atom. This reduces the scaling of the calculation from N^2 to $N \ln N$ while at the same time avoiding the approximations introduced by use of a cutoff.

8.2.4 Molecular model types

There are a number of variations on the classical force field approach described above. One can categorize the different methods by considering the concept of a molecular model. Where a molecular model is defined by the basic elements, such as atoms and bonds, and the nature of the interactions of these elements, described by the force field equation, such as that discussed above, the parameters of the potential and the method used for the representation of solvent. Together these elements form the model. There is a tight connection between a model type and an accompanying force field in the sense that the molecular model and the kinds of behaviors to be modeled determine what needs to be in the force field. However, there can be several force fields, with different parameters and functional forms for the same molecular model type. With this important distinction between the model type and the force field in mind, the most commonly used model types and then the most popular force fields will be discussed, and finally a brief description of current solvent model types. Each of these topics, coupled with the background to classical molecular mechanics described above, deserves a full chapter in itself and so we present only a brief overview of what is most popular and what we believe to be the essential tools in the current state-of-the-art research. For a more in-depth discussion the reader can refer to references Allen and Tildesley (9), Grant and Richards (10), Frenkel and Smit (11), Leach (12), Jensen (13), and Cornell *et al.* (6).

The most detailed and also most used classical mechanical model for biomolecular simulation is the “All-Atom” model in which the basic element of the model is an atom with properties of mass, partial charge, and an atom type which describes its bonding properties and van der Waals parameters. The energy function of a given arrangement of atoms in an

all-atom approach typically conforms to the AMBER force field equation described above which is itself an all-atom model. An all-atom molecular model is created by describing all the atoms and choosing a force field. A full description of atoms includes specifying which atoms are bonded together, what kind of atoms they are, such as an sp^2 carbon, and positioning them in space. The “United Atom” model (14–17) is the same as an all-atom model except that aliphatic hydrogen atoms are combined with the carbon to which they are bonded to form a “united” atom with most of the properties of the carbon, but a larger van der Waals radius and a larger mass. This model requires a different force field than the all-atom model.

Reduced models currently in use are the bead (18) or GO (19, 20) models, models used in large-scale normal mode analysis, and large subunit modeling such as is used in virus capsid assembly modeling (21). Bead models use a whole residue as the element of the model in which each residue, such as an amino acid in a protein, is represented by a single sphere with size and interaction properties, and each bead is bonded to other beads. There is a simple force field for this kind of model, which can include attractions, repulsions, and bonding properties such as angles and dihedrals. It is used primarily for studies of folding of biopolymers. For large-scale normal mode analysis, described later, the Hessian matrix can become impossibly large at $9N^2$ for a system of N atoms. An elastic-network model, in which each alpha carbon of a protein is connected to every other alpha carbon by a spring, reproduces the lowest frequency modes (22–24), which are generally the modes of interest, reducing the size of the problem by at least an order of magnitude. The granularity of the problem has been increased further by combining multiple residues into blocks with only rotational and translational degrees of freedom, bonded together and with an elastic spring network; this model is the RTB model (25). The largest granularity that is worth mentioning is the subunit model used in simulating virus capsid assembly (26), in which each unit represents the basic subunit of a virus, which contains three to nine proteins. Each subunit is a rigid body that interacts with other subunits through interaction points with both attractive and repulsive forces. One could envision using this kind of modeling for interactions between the subunits of the plant cell wall.

8.3 Force fields

From the discussion above it should be obvious that the accuracy of any molecular mechanics method is entirely dependent on the quality of the parameterization used. In theory one could parameterize specifically for each simulation that is to be run but this would be extremely time-consuming and would also make the parameters non-transferable. Instead, in order to preserve the transferability of the force field one attempts to minimize the number of parameters by reducing the system to a set of building blocks. For example, in proteins this is typically the set of individual amino acid residues.

There are a large number of different force fields that have been developed over the years and all have their advantages and disadvantages. For example, the CHARMM force fields include charmm19 (15), charmm22 (27), charmm27 (28), and AMBER force fields include FF94 (6), FF96 (29), FF99 (30), FF99SB (31), FF03 (32, 33). A comprehensive review of the various force fields is beyond the scope of this chapter. For further information the reader is encouraged to read the various papers referring to each force field. In line with the topic

of this book, which is the study of cellulose, we will concentrate our discussion on the force fields available for carbohydrates.

8.3.1 Carbohydrate force fields

The models, above all, require a force field. With the exception of the all-atom model, each model uses a force field that is custom made for the particular molecular model. Here we restrict our discussion to the all-atom force fields, which are the most commonly used and generally considered to be the most accurate. Thus from here on, when we specify a force field, we mean an all-atom force field. As mentioned above there are a number of mature and well established force fields for modeling amino acids and nucleic acids.

The introduction of usable carbohydrate force fields to the AMBER (34, 35), CHARMM (36, 37), OPLS (38, 39), and GROMOS (40) force fields occurred after the force fields were already established and verified for proteins, nucleic acids, and lipids. Carbohydrate research is driven mainly by research in food science, and has largely concentrated on starches and glycosylated amino acids with little attention to cellulose. The small amount of unambiguous experimental data, especially structural data, contributes to the reluctance to model cellulose since the force fields cannot be verified as being appropriate for cellulose modeling in their current forms, and a more complex reparameterization may be necessary to reliably simulate cellulose. The molecular structures associated with cellulose are quite large in terms of current molecular modeling capabilities and require large computational resources. Although the current carbohydrate force fields have been carefully constructed for small molecules and carbohydrates, the force fields have not been tested extensively for such large structures in the same way that the minute details of force fields for proteins have been adjusted to reproduce known structures and known probabilities of alpha helices and beta sheets. Few researchers are willing to apply serious intellectual or computational effort toward such a speculative endeavor. On the other hand, the same lack of unambiguous experimental data empowers modelers to simulate cellulose and its interactions to suggest possible structures and behaviors as well as eliminate highly unlikely ones and to attribute structural features to their underlying physics, even if the modeling is crude. The stage is set for a significant contribution of MD to the understanding of cellulose structure, function, and behavior as new experimental techniques and data are available, and larger and faster computers are accessible to MD modelers.

8.3.2 Solvent models

In the early molecular models, water was not included, but was later determined to be essential to carbohydrate structure and behavior. The solvent model types are classified as explicit solvent, in which solvent molecules are explicitly modeled using the same model type as the solute, or implicit, in which the effect of solvent is modeled by a functional form and is a function of solute configuration. Water is the most common solvent, though other solvents and mixed solvents are also used, especially to reproduce experimental solvent environments. There are several explicit water models, the most important ones being TIP3P, SPC, TIP4P, and TIP5P (41).

The implicit solvent models are based on the assumption that the major effect of solvation is encapsulated in its dielectric properties. In a simple protein of 2400 atoms, solvating with explicit water molecules for a cube with at least three water layers on each side increases the system size to 23 500 atoms. The simplest, and crudest, model simply uses a distance-dependent dielectric constant in the electrostatic term of Equation (8.1), so that the effect of the solvent is to mask the charge interaction between distant charges, assumed to have a dielectric medium between them, and to not mask at all when two charges are close to each other. The advantage of this method is that no explicit water molecules are included in the simulation and the cost of a distance-dependent dielectric constant is minimal, cutting the computational demand by a factor of ten or more. The drawback is that one does not model the solvation free energy correctly nor the dielectric environment inside a macromolecule.

The more sophisticated methods of modeling solvent implicitly are based on solving the Poisson equation. The most rigorous methods involve solving the partial differential equations for the electrostatic potential on a grid, and are quite computationally intensive. These methods, commonly called Poisson–Boltzmann (PB) solvers (42, 43), are useful in accurate examination of electrostatic potentials around static macromolecules and are not often used for dynamics. Even though there are no explicit water molecules in a PB calculation, the computation of the electrostatic potential at each dynamics step is too costly to offset the savings. The Generalized Born approximation is used to provide a much more efficient method for parameterizing the Poisson equation which is very close to the rigorous solution and provides reasonable solvation energies and other thermodynamic solvation effects (44–46). The detailed interactions with individual water molecules are missing as are the hydrodynamic effects, but for many modeling problems, a solvated simulation that reproduces the ensemble averages of an explicitly solvated system can be performed at one fourth the computational cost. A second benefit of implicit solvent calculations is that the solvent response to solute changes is instantaneous at each step, rather than requiring many picoseconds (thousands of steps) of equilibration of the thousands of individual water molecules in an explicit-water simulation. This solvation model is also very useful in preparing a system for fully solvated modeling and for finding probable mechanisms and structures for more detailed studies.

8.4 Molecular dynamics

While using our definition for a classical model above allows us to evaluate the energy of a given arrangement of atoms, and potentially to optimize the structure to a local minimum energy conformation when connected to an optimization algorithm, this approach has only limited scientific value. This is especially true when one wants to study a processive enzyme such as the cellulose hydrolysis enzyme CBH I. In such a situation, it is the dynamics of the system that are of interest to researchers trying to uncover its mode of action with the aim of ultimately improving its efficiency. Thus, to obtain such dynamical properties it is necessary to use the potential energy equation discussed above, and the corresponding gradients (or forces) to propagate the system through time. This is achieved by using dynamics methods that are collectively termed molecular dynamics (MD).

8.4.1 Dynamics methods

The workhorse of the MD methodologies and programs is the dynamics engine that treats the system as a classical mechanical system and integrates Newton's equations of motion based on the force field that is applied. This amounts to initiating some velocities for the atoms, determining forces (the negative gradient of the potential), and then propagating the velocities and adjusting them for the forces one small step at a time. An analytical solution to Newton's equations of motion for even a four-atom molecule using a typical all-atom force field does not exist and thus it is necessary to employ numerical techniques. Numerous numerical algorithms exist for solving the differential equations that arise from Newton's equations of motion, two popular formulations being the predictor-corrector methods (47) and finite difference methods. The most commonly used are the finite difference methods and so the following discussion centers on these methods.

8.4.2 Finite difference methods

Finite difference methods work by splitting the integration into many small steps, each separated by a fixed time δt . At a time t , the force on each atom is calculated from the vector sum of the forces arising from its interactions with the other atoms. From this the acceleration (d^2x_i/dt^2) can be calculated. The accelerations are then combined with the position and velocity data for time t in order to calculate the positions and velocities at a time $t + \delta t$. The iteration is repeated until sufficient time steps have been sampled.

Over the period of the time step the force is assumed to be constant. This places severe constraints on the length of a time step. Ideally the larger time steps allow more phase space to be explored for a given computational effort. However, if the time step is too long the assumption of constant force will break down and lead to instabilities. This causes unrealistic oscillations in the system that can rapidly multiply resulting in an unstable molecular dynamics trajectory. In practice, the time step is limited to an order of magnitude lower than the highest frequency motions. In flexible molecules, these are typically bond stretches involving hydrogen (e.g., C-H ca. 10 fs period). Several methods such as SHAKE (48) and RATTLE (49) exist that, via the use of constraints on bonds with high-frequency oscillations, allow longer time steps to be used. For accurate dynamics, it is generally accepted that only motions involving hydrogen can be constrained and thus the maximum time step is effectively limited to a maximum of 2 fs. This means that to simulate a biological system for 100 ns the complete energy and forces of the entire system have to be evaluated a total of 50×10^9 times in a linear fashion. Since step $n + 1$ cannot be evaluated before step n , it should be immediately apparent that such MD simulations are not applicable to distributed or grid-based computation and instead require very tightly coupled supercomputers with fast processors and extremely low latency and high bandwidth interconnects.

8.4.3 System size limitations

Historically, the need to carry out large numbers of complex calculations in a sequential manner, in order to undertake sufficient sampling to produce meaningful results, has limited the size of a system that could be reasonably modeled to hundreds of atoms in the 1980s,

to millions of atoms now. That limit will continue to be pushed larger but there are new problems associated with systems of even hundreds of thousand atoms. The state-of-the-art computational methods are only now at the point that the smallest cellulosic systems found in nature, such as the plant cell wall cellulose microfibril, can be modeled with confidence. Together with the current force field development, computer hardware technology, and numerical methodologies for high performance computing, the stage is set to probe cellulose and its structures and reactions and answer questions that have been as recalcitrant as the cellulose itself. Reported modeling studies of the cellulose preparations (37, 50–57) are among the few examples of computational structural studies of cellulose. However, Nimlos and coworkers (58) have recently shown that MD simulations of protein–cellulose interactions can shed light on the as yet unknown nature of those interactions. Beyond simple MD simulations, we will discuss the kinds of numerical simulations and the properties that can be studied, quantified, and predicted.

8.4.4 Quantum mechanics/molecular dynamics

As mentioned above, one of the biggest limitations of classical MD simulations is the inability to make and break covalent bonds during a simulation. A potential solution to this is to describe the system quantum mechanically instead of classically. This would seem to be an ideal situation since then electron densities are explicitly included in the calculation. However, such simulations are far too computationally expensive to be used for molecular dynamics simulations of proteins. An alternative approach, first proposed in 1976 by Warshel and Levitt (59), was to combine a quantum mechanical (QM) potential with a molecular mechanical (MM) potential to form a hybrid QM/MM potential. In this approach, the parts of the protein and substrate that are directly involved in the enzyme reaction are calculated using QM potential functions, and the remaining atoms are treated using a classical MM potential. The coupling of a QM and MM potential allows just the reaction center to be studied quantum mechanically while keeping the calculation complexity low by using a more approximate MM potential elsewhere. This partitioning of the system allows calculations on systems significantly larger than would be possible with pure QM approaches and at the same time enables calculations such as reactions to be studied for which classical MM potentials are not appropriate. A number of commonly used MD codes provide support for QM/MM simulations including AMBER and CHARMM. QM/MM simulations still remain relatively expensive, however, although recent advances in AMBER (60) are bringing the cost of QM/MM simulations, for systems with up to 100 QM atoms, to a point where the cost is approximately double what the standard classical simulation would cost. Thus, QM/MM approaches will form an important tool for looking at the actual mechanism of hydrolysis within a cellulase enzyme as it degrades cellulose fibers.

8.5 Analysis methods

The previous section has provided a very basic introduction to the concept of molecular dynamics. However, up to this point we have only covered what is required to run a simulation and obtain the variation in atomic positions as a function of time. The sobering fact is that even with computers many orders of magnitude faster than today's machines we will still

not be able to achieve biological time scales for all but the very fastest enzymes. Thus, there is little to be gained scientifically in just running an MD simulation and waiting to see the system react. Firstly, the force field prevents covalent reactions from occurring and even if a method that supports bond breaking is used, such as QM/MM, the simulation time scales are too short for direct observation. Thus, it is necessary to make use of a large number of different techniques and analysis methods in order to uncover scientifically meaningful information. There are a huge number of different analysis techniques that can be used with MD methods and more are being devised every day. Here we list some of the more common and useful techniques for studying cellulose and cellulose hydrolysis but this represents a very small subset of what is possible.

Clustering. Clustering is a method of postprocessing the sampling results of an MD simulation to quantify the number of more highly populated states the system visits, and can be used to reprocess a dynamics trajectory to map the essential transitions between states. In particular, when multiple structural states are observed in MD simulations, the identity of the states and the populations in the MD simulation are quantified in a cluster analysis. Once one knows the states, the sequence of transitions between states can be reconstructed from the same trajectory, revealing the nature and frequency of transitions and the paths through intermediate states if there are any. Cluster analysis can be used to analyze any trajectory, whether it is constrained in any way or not, so care must be taken to interpret the results within the framework of any biasing potentials. In cellulose modeling, this kind of analysis is essential for characterizing crystalline and amorphous states and the transitions between them. For example, it is possible to use cluster analysis to test the hypothesis that a system goes straight from state A to state B, the transition being observed in either experiment or simulation. The analysis may show instead, that there is a highly populated state C that the system always, in simulation, goes through on the way, and never goes straight from A to B or B to A.

Normal mode analysis. Normal mode analysis is a powerful tool for extracting the characteristic motions of a macromolecule or complex such as a cellulose fiber, cellulose/lignin complex or protein/cellulose system. With this method, the high-frequency modes can be separated out allowing the low frequency and larger displacement motions to be identified. Often, those are the motions that define the behavior and the biological function of macromolecules. The method involves finding a structure of the complex which is at an energy minimum and subsequent diagonalization of the mass-weighted Hessian matrix from which the frequencies of motion and normal mode vector (eigenvectors) can be extracted. However, for larger molecules, the diagonalization of the matrix can quickly become too large a problem for most computers. There are other methods of treating larger molecules and systems of molecules that rely on simple assumptions. One method of treating the larger systems is to minimize the structure to a minimum energy state using the all-atom model, then switch a more approximate method such as the elastic network model, where selected atoms such as alpha carbons are all connected together by a series of harmonic springs, or the Rotations Translations of Blocks (RTB) model which uses an approximate diagonalization method by combining multiple residues into rigid blocks, and then apply the normal mode method (25). Quasiharmonics is a variation of the normal mode analysis in which the effective modes of vibration are calculated from fluctuations which are determined from a MD simulation. Since the fluctuations in an MD simulation contain anharmonic contributions, the quasiharmonic vibrational

modes may differ from the normal modes calculated from the energy minima. The normal mode method has been extended to follow troughs in the potential energy surface, and give information about essential modes of extrema besides minima such as saddles, or transition states, and maxima.

Local water density. Local water density concerns are important when considering many biological models, especially in cases where complex structures compartmentalize regions of water away from bulk water. Considering the complexity of these situations, readers are encouraged to consult specific examples in the literature.

8.6 Enhanced sampling and free energy methods

In addition to using analysis methods to look for scientifically relevant information in molecular dynamics trajectories it is also possible to increase the sampling rate, or force a reaction event to occur. Such approaches are termed enhanced sampling methods and if used carefully can be used to test a number of hypotheses about reaction mechanisms.

Restraints. Restraining potentials are used to keep a system within a defined region of configuration space. Most restraints take the form of a harmonic potential placed on some defined quantity such as distance between two atoms, distance between two parts of a molecule, radius of gyration, or root mean square distance (RMSD) from some reference state. When restraints are used, the states of the system are sampled for configurations that are near the minimum of the restraining potential. NOE distances, from Nuclear Magnetic Resonance experiments, can be used as restraints to refine structures by exploring configurations but keeping within the defined NMR structural information. Structural changes can be followed or initiated by slowly changing restraining potentials from an initial state to a final state, thus achieving a change that might never occur with a simple MD simulation. Ensemble averages can be extracted from the biased runs to calculate thermodynamic properties, so simulations using restraints are not only exploratory but also for collecting data in state space that are rarely or never visited in normal MD runs.

Steered MD. This term refers to forcing MD simulations to follow a trajectory that is biased by a force or biasing potential much like a restraint but with many variations and not necessarily related to a well-defined potential. One clear example is modeling an Atomic Force Microscopy experiment using either a constant pulling force or constant pulling velocity where one end of a molecule is fixed while the other end is pulled in a particular direction. Another example is targeted MD (TMD) (61) in which a final structure is the target and a force is applied to the system, such as a force on the RMSD from the final structure or the Euclidean distance of some atoms from some target, until the target is reached. These methods rarely provide accurate or precise energetic or thermodynamic data, but the approximate data is very useful for probing unknown pathways and providing insight into designing more accurate simulations of the processes or structures of interest. The method of Jarzynski (62, 63) can be used with these methods to gain ensemble statistics if a large number of pulling or targeted simulations are performed, but often this method is no less computationally demanding than the slower sampling methods due to the sheer number of trajectories that must be run. In a very practical sense, this method is very useful to bring a system to a structural state of interest, for which there

is no crystal or NMR structure, such as docking a ligand into a site allowing the nearly natural reconfiguration of the receptor in the process.

Nudged elastic band. The nudged elastic band (NEB) method provides a method for locating low energy transition pathways in biological systems. In NEB (64, 65), the minimum energy path for a conformational change is quantified with a series of images of the molecule describing the path. The images at the endpoints remain fixed in space while each image in-between is connected to its immediate neighbors by “springs” along the pathway that act to keep each image from sliding down the energy landscape onto adjacent images. By running simulated annealing, followed by quenched MD, it is possible to freeze out these images equally spaced along a low energy reaction pathway. The advantage of the NEB method over traditional transition path sampling calculations is that it does not require an initial hypothesis for the pathway. Such simulations, when coupled with QM/MM approaches to allow for bond breaking and formation, will likely prove extremely useful in studying the catalytic action of cellulases acting on cellulose substrates. In particular, such approaches may offer insights into the processive nature of such enzymes.

8.6.1 Free energy methods

8.6.1.1 Thermodynamic integration

Thermodynamic integration is the method of obtaining a Helmholtz free energy change, ΔF , between two states where the difference is determined from averages that are accessible from MD simulations. If a path between two states can be defined and states along that path can be defined by a parameter, $\lambda \in (\lambda_0, \lambda_1)$, then the following statistical mechanical relationship can be derived,

$$\frac{dF_\lambda}{d\lambda} = \left\langle \frac{\partial U_\lambda}{\partial \lambda} \right\rangle_\lambda \quad (8.6)$$

where U_λ is the potential energy at the state defined by λ , and the average is over all configurations visited by the system during an MD simulation of the system with constant λ . Then using simple integration schemes, integrating from initial state λ_0 to final state λ_1 , for simplicity we choose $\lambda_0 = 0$ and $\lambda_1 = 1$,

$$\Delta F = F_{\lambda=1} - F_{\lambda=0} = \int_0^1 d\lambda \left\langle \frac{\partial U_\lambda}{\partial \lambda} \right\rangle_\lambda \quad (8.7)$$

Although there are some very delicate issues regarding the endpoints of this method, the beauty is that one can run a few points between each of the endpoints, and using Gaussian quadrature, integrate to high precision the averages of the potential energy derivatives to yield the ΔF . In addition to the simplicity of the method, the ability to choose any path from state 0 to state 1 including creating and annihilating atoms. A typical application is to use a thermodynamic cycle to find the change in binding free energy, $\Delta \Delta F_{\text{binding}}$, for two ligands binding to a substrate without calculating the binding free energy of either one. Free energies of binding are particularly hard to calculate. The method can be illustrated by the following free energy cycle, whose total free energy change will be zero since the beginning and ending states are the same state. In this illustration, Figure 8.4, the two ligands are

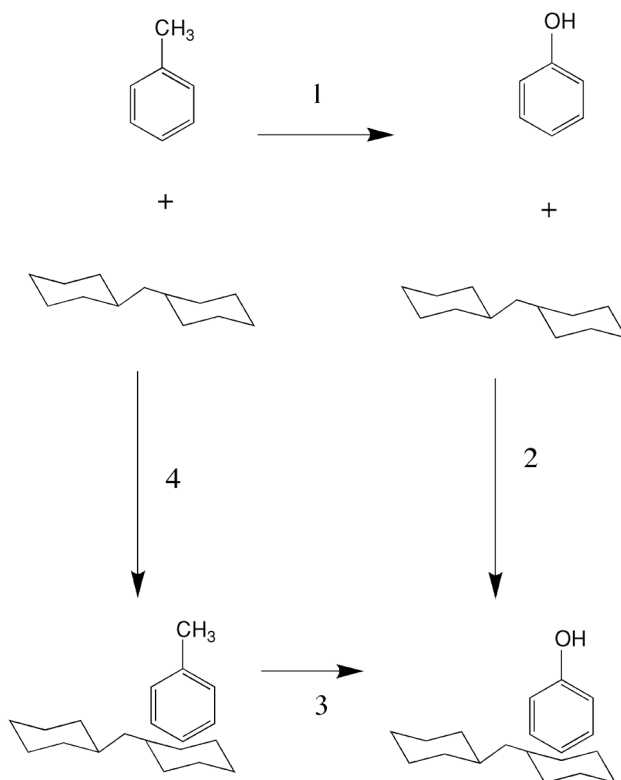


Figure 8.4 A thermodynamic cycle used to determine the difference in free energy of binding of toluene and phenol to cellulose.

phenol and toluene, which differ by the constituent group being hydroxyl or methyl. The question to answer is which binds to cellulose [100] surface more strongly, as judged by the difference in free energy of binding, or $\Delta\Delta F_{\text{binding}}$, or the difference in ΔF between the binding processes, 4 and 2. The free energy change for the sum of reactions 1 and 2 must be equal to the free energy change for the sum of reactions 4 and 3 since they have the same initial and final states. This relationship is expressed, for the reactions going in the directions indicated in the figure,

$$\Delta F_4 + \Delta F_3 = \Delta F_1 + \Delta F_2 \quad (8.8)$$

which can be rearranged to yield the difference in binding free energies desired,

$$\Delta F_4 - \Delta F_2 = \Delta F_1 - \Delta F_3 = \Delta\Delta F_{\text{binding}} \quad (8.9)$$

With this relation, we can determine the difference between two processes that are very hard to compute using two processes that are straightforward and relatively easy. In particular, both processes involve only changing CH3 to OH, and process 1 does not even involve the cellulose since there is no change in its solvated state in process 1. Calculation of the

significant entropic contribution to the binding process is eliminated in this method since the difference of the two binding processes largely cancels out the entropic contributions. The simulations do not have to correctly simulate the large changes associated with desolvating both the cellulose and the ligand, only the small solvation and structural entropy changes associated with changing a small functional group.

8.6.1.2 Potential of mean force and umbrella sampling

The potential of mean force (PMF) is used to get information about how the free energy of a system changes as it moves along a particular trajectory, which could represent a conformational change or chemical reaction. The key relationship is

$$P(\mathbf{x}) \propto e^{-A(\mathbf{x})/k_b T} \quad (8.10)$$

where $P(\mathbf{x})$ is the probability of state \mathbf{x} , $A(\mathbf{x})$ is the Helmholtz free energy of state \mathbf{x} , k_b is Boltzman's constant, and T is the temperature. State \mathbf{x} is defined as a configuration or reaction coordinate and is constrained to a particular place in configuration space while the rest of the system is averaged over all accessible states. The PMF is useful in following the change in Helmholtz free energy as the system moves along the trajectory coordinate, most often to find the change in free energy between initial state, the transition and final state. In some cases, this probability can be arrived at by a simple MD simulation, run for sufficiently long time that all states \mathbf{x} are sampled frequently enough to yield valid probabilities. The method of straight MD breaks down for this method when the barriers are high enough that the states around the transition state do not get visited even for very long simulations. The answer to this problem is to use some method of enhanced sampling, such as umbrella sampling (66). Umbrella sampling is implemented by applying a biasing potential, usually harmonic similar in shape to an inverse umbrella, centered at various points along the defined trajectory or reaction path and constructed such that the umbrellas are large enough and the points are close enough that the sampling within each umbrella overlaps with the adjacent umbrella sampling. The system will sample within the umbrella potential and give information about the probabilities of states within the umbrella, though they are biased. These probabilities can be unbiased but still will contain an unknown constant. The pieces can be spliced directly back together lining up the overlap regions; more commonly the more exact method, weighted histogram analysis method (WHAM) (67), is used to produce the full PMF for the trajectory. Other methods, not discussed here, which are useful for the same problems as umbrella sampling and can often be used in cases where umbrella sampling does not work, are: locally enhanced sampling (68), replica exchange (69), and lambda dynamics (70).

8.7 Studying cellulose hydrolysis

8.7.1 Work to date

The questions researchers have tried to answer over the last two decades using molecular mechanics and molecular dynamics have been, what are the structures of cellulose isotypes I_α , I_β , and II , and, can the uncertainties and irregularities in the experimental data be

clarified? Many insights have been gained from simulations using multiple force fields, models, and programs. The structure of I_α has been robust in its behavior across several of the force fields as evidenced by the average structures after sufficient equilibration (37, 50–54, 56, 57, 71, 72). These simulations have been able to show the details of the tilting of sugar rings in alternating layers, the right-handed twist in the cellulose fiber, and the stable hydrogen bonding and changes in hydrogen bonding (53) over nanosecond time scales. Analysis of the water structure (57, 73, 74) and simulations of smaller bundles of celloextrin chains suggest that the structure is more gel-like (72) in contrast to the simulations of larger bundles in which highly crystalline structure is found to be stable below the first layer. The study of the free energy surface of the α -(1-4)-glycosidic linkage including the effects of water (75) provides insight into the contribution of this linkage to more complex cellulosic structures using umbrella sampling to produce the surface.

More recently, MD tools have been used to study interactions between cellulose and other non-cellulosic molecules. The binding domains of cellulases have been docked onto the [100] surface of cellulose I_α resulting in observation of interesting and possibly important behavior in multi-nanosecond MD simulations (58). The entire CBH I enzyme was docked onto the cellulose surface and simulated for 1.5 ns. The complex was found to be stable and have significant water structuring around the interaction regions (76).

8.7.2 Approaches to current questions about structure and hydrolysis

There are two major approaches to hydrolysis of cellulose, acid hydrolysis, and enzymatic hydrolysis. The enzymatic process is poorly understood and must contain the solution to the recalcitrant nature of cellulosic degradation. The enzymes can be modeled, as well as their interactions with cellulose and even the process of enzymatic hydrolysis. The techniques that will probe the processes and mechanisms are numerous and range from reduced models to all-atom QM/MM and thermodynamic integration. Using reduced models, the structural stabilities and solvation free energies can be determined quickly. Normal mode and elastic network models and quasiharmonic analysis can probe the major structural modes of motion of cellulose, cellulases, xylans, lignins, and their mutual complexes. Mutational studies, using thermodynamic integration, can be performed to reveal the effects on structure and on kinetic behaviors, and even on reaction energetics and mechanisms. Umbrella sampling is a key player in understanding the binding affinities of different binding or catalytic domains on cellulose, or the relative binding affinities on different faces of cellulose or even on different locations of the same face. QM/MM is a tool for probing the hydrolysis reaction inside a cellulase catalytic site. This method is at the stage of development that performance is sufficient and the QM approximations are good enough to follow a reaction quantum mechanically while treating the non-reactive portion of the system classically and have reasonable answers for not much higher computational cost than pure classical simulations. It is expected that exceptionally useful information about the release of energy from reaction, and the accompanying structural changes will come from these numerical experiments.

The steered molecular dynamics, targeted MD, and pulling methods are the tools of choice for initial examination of the process of decrystallization of the cellulose fibers into celloextrin chains suitable for hydrolysis to mono- and disaccharides. These kinds of numerical experiments can suggest the energy barriers associated with decrystallization, and suggest

more detailed studies such as obtaining PMF profiles from umbrella sampling runs, or free energies of decrystallization from Jarzynski pulling experiments. Beyond that, details about how the solvent plays a role in all the aforementioned processes can be carefully quantified and help to select the most likely and deselect unlikely mechanisms.

8.8 Performance and future of cellulose modeling

A complete discussion of all the issues surrounding the performance of molecular dynamics simulations is beyond the scope of this book and remains an active area of research in its own right. However, this chapter would not be complete without at least some discussion of the typical resource requirements for MD simulations of celluloses and the types of time scales that can be reached.

MD simulations are very computationally intense, typically requiring access to the world's most powerful supercomputers in order to simulate sufficient timescales (10 ns to 1 μ s) for biological systems (10 thousand to several million atoms). The computational complexity lies in the fact that there are an extremely large number of pairwise interactions that must be calculated at each time step. This would typically be on the order of 20 million or more for a system of 80 000 atoms. Then to access information on a biological timescale, it is necessary to propagate the system through time. As mentioned above, the length of a time step is typically limited to about 2 fs if the motions of all heavy atoms are to be simulated. This means that to obtain 100 ns of trajectory data requires evaluation of 5×10^{10} time steps and each time step requires the 20 million non-bond energy evaluations.

Unlike Monte Carlo simulations, where each individual energy evaluation can be performed independently of other calculations, molecular dynamics simulations involve numerically solving the integral over time. This means that the next step of the MD trajectory cannot be computed until all previous steps have been computed in order. This makes computation in parallel difficult, requiring extremely low latency interconnects between processors and careful distribution of work. Even then, the distribution of work across multiple CPUs is limited to a single time step. There is a multiple time step method in which the low-frequency motions are partially uncoupled from the high-frequency motions in such a way that low-frequency contributions to the dynamics are not calculated on every step, increasing performance as much as a factor of two. There is significant effort being expended in developing more efficient molecular dynamics software.

8.8.1 Current performance

There are a large number of molecular dynamics packages currently available and, while performance varies across each piece of software and the type of method and approximations employed, they all effectively simulate on the nanosecond timescales out to limits of tens to hundreds of nanoseconds. In the next few years this will likely be extended into the microsecond regime, but we are unlikely to see any form of routine simulation above several microseconds of simulation time for many years to come. It should be noted that simply extending the length of time for which a system is simulated will only go so far toward improving the results. For example, to simulate rare events, it will always be necessary to

carry out more complex sampling using some of the methods described above. Thus, it will be essential in the coming years that effort is focused not just on improving the performance of MD codes for simple linear MD simulations, but also on all the advanced sampling approaches that have been developed.

To give some context to the types of resource requirements simulations of cellulose require, a simulation of a cellulose bundle in explicit water comprising 408 000 atoms runs at a rate of 2.7 ns/day on 768 processors of San Diego Supercomputer Center's DataStar (IBM Power 4) supercomputer using version 9 of Amber's PMEMD molecular dynamics software. Thus, a 20 ns long simulation requires access to such a supercomputer for 7.4 days and utilizes a total of approximately 136 000 CPU hours.

8.9 Future possibilities

When the large computers of the future, combined with the advances in the computational tools arrive, we can expect to see all-atom models studying millions of atom-sized structures and producing statistical-thermodynamic results based on a sampling scale in the microsecond range. Modeling techniques will grow into the multi-scale modeling arena such that large-scale structures can be modeled reliably including the cell wall and the large structures associated with it. Figure 8.5 depicts one potential approach to developing a programmatic solution for modeling the plant cell wall, starting with the simplest problems addressable today (e.g., still a challenge today) and building toward computational interpretations of the

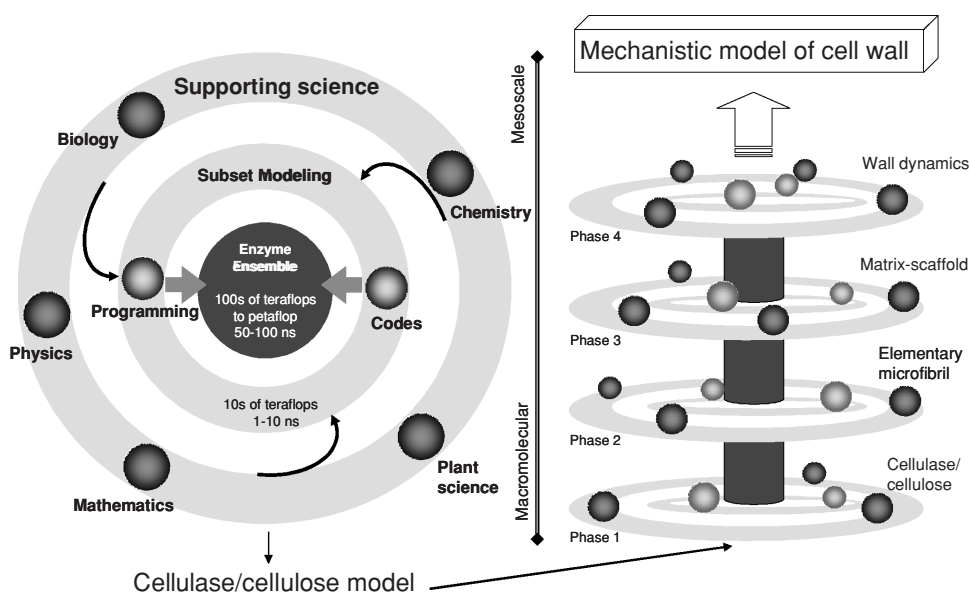


Figure 8.5 Depiction of a multi-scale approach to modeling the plant cell wall where the problem is parsed in subsets solvable in near term, followed by combinations of subsets into ensemble problems studied as advanced codes and computers become available. (provided by Michael Himmel)

wall system using codes and processors not yet available. Computational approaches to the cellulose hydrolysis problem will only become more vital to the full understanding of the processes involved as the computational architectures and methods evolve, working harmoniously with experimental and theoretical approaches to solve what are now unapproachable problems.

Acknowledgments

This work was supported by the DOE Office of the Biomass Program and by the NSF via the Strategic Applications Collaboration program at the San Diego Supercomputer Center.

References

1. Born, M. & Oppenheimer, R. (1927) Zur Quantentheorie Der Molekeln. *Annals of Physics*, **84**, 457–484.
2. Debolt, S.E. & Kollman, P.A. (1990) A theoretical-examination of solvatochromism and solute solvent structuring in simple alkyl carbonyl-compounds – simulations using statistical mechanical free-energy perturbation-methods. *Journal of the American Chemical Society*, **112**, 7515–7524.
3. Allinger, N.L. (1977) Conformational-analysis. 130. MM2 – hydrocarbon force-field utilizing V1 and V2 torsional terms. *Journal of the American Chemical Society*, **99**, 8127–8134.
4. Allinger, N.L., Yuh, Y.H. & Lii, J.H. (1989) Molecular mechanics – the MM3 force-field for hydrocarbons. 1. *Journal of the American Chemical Society*, **111**, 8551–8566.
5. Mackerell, A.D., Wiorkiewicz-Kuczera, J. & Karplus, M. (1995) An all-atom empirical energy function for the simulation of nucleic-acids. *Journal of the American Chemical Society*, **117**, 11946–11975.
6. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. & Kollman, P.A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *Journal of the American Chemical Society*, **117**, 5179–5197.
7. Hermans, J., Berendsen, H.J.C., Vangunsteren, W.F. & Postma, J.P.M. (1984) A consistent empirical potential for water–protein interactions. *Biopolymers*, **23**, 1513–1518.
8. Darden, T., Perera, L., Li, L.P. & Pedersen, L. (1999) New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure with Folding and Design*, **7**, R55–R60.
9. Allen, M.P. & Tildesley, D.J. (1987) *Computer Simulation of Liquids*. Oxford Science Publications, New York.
10. Grant, G.H. & Richards, W.G. (1995) *Computational Chemistry*. Oxford University Press, Oxford.
11. Frenkel, D. & Smit, B. (1996) *Understanding Molecular Simulation*. Academic Press, San Diego.
12. Leach, A.R. (2001) *Molecular Modelling: Principles and Applications*. Prentice Hall, Englewood Cliffs.
13. Jensen, F. (2006) *Introduction to Computational Chemistry*. Wiley, New York.
14. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984) A new force-field for molecular mechanical simulation of nucleic-acids and proteins. *Journal of the American Chemical Society*, **106**, 765–784.
15. Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983) Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, **4**, 187–217.

16. Smondyrev, A.M. & Berkowitz, M.L. (1999) United atom force field for phospholipid membranes: Constant pressure molecular dynamics simulation of dipalmitoylphosphatidicholine/water system. *Journal of Computational Chemistry*, **20**, 531–545.
17. Yang, L.J., Tan, C.H., Hsieh, M.J., Wang, J.M., Duan, Y., Cieplak, P., Caldwell, J., Kollman, P.A. & Luo, R. (2006) New-generation amber united-atom force field. *Journal of Physical Chemistry B*, **110**, 13166–13176.
18. Tozzini, V., Rocchia, W. & McCammon, J.A. (2006) Mapping all-atom models onto one-bead coarse-grained models: General properties and applications to a minimal polypeptide model. *Journal of Chemical Theory and Computation*, **2**, 667–673.
19. Karanicolas, J. & Brooks, C.L. (2003) Improved go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *Journal of Molecular Biology*, **334**, 309–325.
20. Shea, J.-E., Onuchic, J.N. & Brooks, C.L., III. (1999) Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A. *PNAS*, **96**, 12512–12517.
21. Rapaport, D.C. (2004) Self-assembly of polyhedral shells: A molecular dynamics study. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, **70**, 051905.
22. Bahar, I. & Rader, A.J. (2005) Coarse-grained normal mode analysis in structural biology. *Current Opinion in Structural Biology*, **15**, 586–592.
23. Delarue, M. & Sanejouand, Y.H. (2002) Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: The elastic network model. *Journal of Molecular Biology*, **320**, 1011–1024.
24. Suhre, K. & Sanejouand, Y.H. (2004) On the potential of normal-mode analysis for solving difficult molecular-replacement problems. *Acta Crystallographica Section D – Biological Crystallography*, **60**, 796–799.
25. Tama, F., Gadea, F.X., Marques, O. & Sanejouand, Y.H. (2000) Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins – Structure Function and Genetics*, **41**, 1–7.
26. Kim, M.K., Jernigan, R.L. & Chirikjian, G.S. (2003) An elastic network model of Hk97 capsid maturation. *Journal of Structural Biology*, **143**, 107–117.
27. MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*, **102**, 3586–3616.
28. Mackerell, A.D., Jr, Feig, M. & Brooks, C.L. III. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*, **25**, 1400–1415.
29. Kollman, P.A., Dixon, R., Cornell, W., Fox, T., Chipot, C. & Pohorille, A. (1997) The development/application of a “Minimalist” organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In: *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Vol. 3 (eds. W.F. Van Gunsteren, P.K. Weiner & A. Wilkinson), pp. 83–96. Escom, Leiden, The Netherlands.
30. Wang, J.M., Cieplak, P. & Kollman, P.A. (2000) How well does a restrained electrostatic potential (Resp) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, **21**, 1049–1074.
31. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins – Structure Function and Bioinformatics*, **65**, 712–725.

32. Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G.M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J.M. & Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, **24**, 1999–2012.
33. Lee, M.C. & Duan, Y. (2004) Distinguish protein decoys by using a scoring function based on a new amber force field, short molecular dynamics simulations, and the generalized born solvent model. *Proteins – Structure Function and Bioinformatics*, **55**, 620–634.
34. Woods, R.J., Dwek, R.A., Edge, C.J. & Fraser-Reid, B. (1995) Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. Glycam_93 parameter development. *Journal of Physical Chemistry*, **99**, 3832–3846.
35. Kirschner, K.N. & Woods, R.J. (2001) Solvent interactions determine carbohydrate conformation. *Proceedings of the National Academy of Sciences*, **98**, 10541–10545.
36. Kuttel, M., Brady, J.W. & Naidoo, K.J. (2002) Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *Journal of Computational Chemistry*, **23**, 1236–1243.
37. Reiling, S. & Brickmann, J. (1995) Theoretical investigations on the structure and physical properties of cellulose. *Macromolecular Theory and Simulations*, **4**, 725–743.
38. Kony, D., Damm, W., Stoll, S. & van Gunsteren, W.F. (2002) An improved OPLS-AA force field for carbohydrates. *Journal of Computational Chemistry*, **23**, 1416–1429.
39. Damm, W., Frontera, A., TiradoRives, J. & Jorgensen, W.L. (1997) OPLS all-atom force field for carbohydrates. *Journal of Computational Chemistry*, **18**, 1955–1970.
40. Ott, K.H. & Meyer, B. (1996) Parametrization of GROMOS force field for oligosaccharides and assessment of efficiency of molecular dynamics simulations. *Journal of Computational Chemistry*, **17**, 1068–1084.
41. Mahoney, M.W. & Jorgensen, W.L. (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *Journal of Chemical Physics*, **112**, 8910–8922.
42. Baker, N.A. (2005) Improving implicit solvent simulations: A Poisson-centric view. *Current Opinion in Structural Biology*, **15**, 137–143.
43. Cramer, C.J. & Truhlar, D.G. (1999) Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chemical Reviews*, **99**, 2161–2200.
44. Onufriev, A., Bashford, D. & Case, D.A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins – Structure Function and Bioinformatics*, **55**, 383–394.
45. Feig, M. & Brooks, C.L. (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Current Opinion in Structural Biology*, **14**, 217–224.
46. Mongan, J., Simmerling, C., McCammon, J.A., Case, D.A. & Onufriev, A. (2007) Generalized born model with a simple, robust molecular volume correction. *Journal of Chemical Theory and Computation*, **3**, 156–169.
47. Gear, C. (1971) *Numerical Initial Value Problems in Ordinary Differential Equations*. Prentice-Hall, Englewood Cliffs.
48. Ryckaert, J.P., Ciccotti, G. & Berendsen, H.J.C. (1977) Numerical-integration of cartesian equations of motion of a system with constraints – molecular-dynamics of N-alkanes. *Journal of Computational Physics*, **23**, 327–341.
49. Andersen, H.C. (1983) Rattle – a velocity version of the shake algorithm for molecular-dynamics calculations. *Journal of Computational Physics*, **52**, 24–34.
50. Alvo Aabloo, A.D.F. (1994) Preliminary potential energy calculations of cellulose Ia crystal structure. *Macromolecular Theory and Simulations*, **3**, 185–191.
51. Heiner, A.P., Sugiyama, J. & Teleman, O. (1995) Crystalline cellulose I[alpha] and I[beta] studied by molecular dynamics simulation. *Carbohydrate Research*, **273**, 207–223.

52. Kroon-Batenburg, L.M.J. & Kroon, J. (1997) The crystal and molecular structures of cellulose I and II. *Glycoconjugate Journal*, **14**, 677–690.
53. Chen, W., Lickfield, G.C. & Yang, C.Q. (2004) Molecular modeling of cellulose in amorphous state part II: Effects of rigid and flexible crosslinks on cellulose. *Polymer*, **45**, 7357–7365.
54. Mazeau, K. (2005) Structural micro-heterogeneities of crystalline I β -cellulose. *Cellulose*, **12**, 339–349.
55. Yui, T., Nishimura, S., Akiba, S. & Hayashi, S. (2006) Swelling behavior of the cellulose I beta crystal models by molecular dynamics. *Carbohydrate Research*, **341**, 2521–2530.
56. Yui, T. & Hayashi, S. (2007) Molecular dynamics simulations of solvated crystal models of cellulose I α and III β . *Biomacromolecules*, **8**, 817–824.
57. Matthews, J.F., Skopec, C.E., Mason, P.E., Zuccato, P., Torget, R.W., Sugiyama, J., Himmel, M.E. & Brady, J.W. (2006) Computer simulation studies of microcrystalline cellulose I beta. *Carbohydrate Research*, **341**, 138–152.
58. Nimlos, M.R., Matthews, J.F., Crowley, M.F., Walker, R.C., Chukkappalli, G., Brady, J.W., Adney, W.S., Cleary, J.M., Zhong, L. & Himmel, M.E. (2007) Molecular modeling suggests induced fit of family I carbohydrate-binding modules with a broken-chain cellulose surface. *Protein Engineering, Design and Selection*, **20**, 179–187.
59. Warshel, A. & Levitt, M. (1976) Theoretical studies of enzymic reactions – dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *Journal of Molecular Biology*, **103**, 227–249.
60. Walker, R.C., Crowley, M.F. & Case, D.A. (2007) The Implementation of a fast and accurate QM/MM potential method in Amber. *Journal of Computational Chemistry*, in press.
61. Ma, J., Sigler, P.B., Xu, Z. & Karplus, M. (2000) A dynamic model for the allosteric mechanism of Groel. *Journal of Molecular Biology*, **302**, 303–313.
62. Gore, J., Ritort, F. & Bustamante, C. (2003) Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 12564–12569.
63. Fox, R.F. (2003) Using nonequilibrium measurements to determine macromolecule free-energy differences. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 12537–12538.
64. Jonsson, H., Mills, G. & Jacobsen, K.W. (1998) Nudged elastic band method for finding minimum energy paths of transitions. In: *Classical and Quantum Dynamics in Condensed Phase Simulations* (eds. B.J. Berne, G. Ciccoli & D.F. Coker), pp. 385–404. World Scientific, Singapore.
65. Mills, G. & Jonsson, H. (1994) Quantum and thermal effects in H-2 dissociative adsorption – evaluation of free-energy barriers in multidimensional quantum-systems. *Physical Review Letters*, **72**, 1124–1127.
66. Torrie, G.M. & Valleau, J.P. (1977) Non-physical sampling distributions in Monte-Carlo free-energy estimation – umbrella sampling. *Journal of Computational Physics*, **23**, 187–199.
67. Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A. & Rosenberg, J.M. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *Journal of Computational Chemistry*, **13**, 1011–1021.
68. Roitberg, A. & Elber, R. (1991) Modeling side-chains in peptides and proteins – application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *Journal of Chemical Physics*, **95**, 9277–9287.
69. Sugita, Y. & Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, **314**, 141–151.
70. Banba, S., Guo, Z.Y. & Brooks, C.L. (2000) Efficient sampling of ligand orientations and conformations in free energy calculations using the lambda-dynamics method. *Journal of Physical Chemistry B*, **104**, 6903–6910.

71. Hardy, B.J. & Sarko, A. (1996) Molecular dynamics simulations and diffraction-based analysis of the native cellulose fibre: Structural modelling of the I-alpha and I-beta phases and their interconversion. *Polymer*, **37**, 1833–1839.
72. Tanaka, F. & Okamura, K. (2005) Characterization of cellulose molecules in bio-system studied by modeling methods. *Cellulose*, **12**, 243–252.
73. Heiner, A.P., Kuutti, L. & Teleman, O. (1998) Comparison of the interface between water and four surfaces of native crystalline cellulose by molecular dynamics simulations. *Carbohydrate Research*, **306**, 205–220.
74. Heiner, A.P. & Teleman, O. (1997) Interface between monoclinic crystalline cellulose and water: Breakdown of the odd/even duplicity. *Langmuir*, **13**, 511–518.
75. Kuttel, M.M. & Naidoo, K.J. (2005) Free energy surfaces for the alpha(1-4)-glycosidic linkage: Implications for polysaccharide solution structure and dynamics. *Journal of Physical Chemistry B*, **109**, 7468–7474.
76. Zhong, L., Matthews, J.F., Crowley, M.F., Rignall, T., Talón, C., Cleary, J.M., Walker, R.C., Chukkappalli, G., McCabe, C., Nimlos, M.R., Brooks, C.L., III, Himmel, M.E. & Brady, J.W. (2008) Interactions of the complete cellobiohydrolase I from *Trichodera reesei* with microcrystalline cellulose I β . *Cellulose*, published online. DOI No. 10.1007/s10570-007-9186-0.

Chapter 9

Mechanisms of Xylose and Xylo-oligomer Degradation During Acid Pretreatment

Xianghong Qian and Mark R. Nimlos

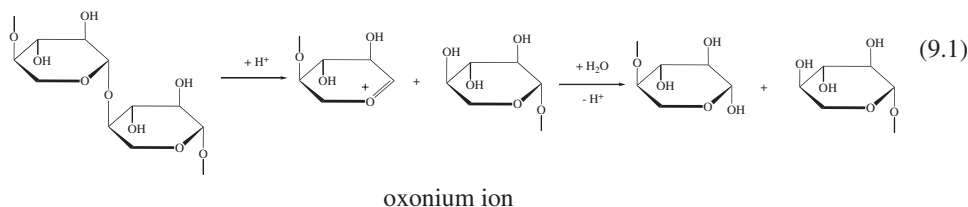
9.1 Background

With the depletion of fossil fuels and the increase in oil prices, biofuels have become ever more attractive as an alternative source of energy. The US Department of Energy (DOE) recently released its roadmap aiming to reach the goal of supplying 30% of the US motor vehicle gasoline from cellulosic ethanol (1). In addition, the European Union has a plan to produce 25% of its transportation fuels from biofuels (2). Toward these goals, significant progress has been made in biomass conversion of cellulose to fermentable sugars. Many schemes for the utilization of biomass as a source of renewable fuels and chemicals rely upon the ability to deconstruct the polysaccharides in plant cell walls into constituent sugars.

Acid hydrolysis is commonly used during pretreatment to break down the structures of plant cell walls and prepare plant polysaccharides for hydrolysis using biological catalysts. In this step, hemicellulose is hydrolyzed to monomeric sugars, the majority of which are pentosans, such as β -D-xylose. Depending on the severity (temperature and acidity) of this pretreatment process, some xylose molecules undergo an undesirable dehydration process, thus lowering the biomass conversion efficiency. In addition, the main degradation product, furfural, is a toxin for fermentative organisms (3, 4) and can polymerize to reduce access to other polysaccharides, such as cellulose. Thus, dehydration reactions in acid pretreatment of biomass present a barrier to economical conversion of biomass. In the past, sugar yield and acidic sugar degradation products were found to be strongly dependent upon the reactor configuration, the reaction media, and the reaction temperature (5–9) during dilute acid hydrolysis.

Xylan is a prevalent hemicellulose in many sources of biomass, and for example, makes up roughly 20% of corn stover (10). The polymer backbone of xylan is composed of xylose monomers joined by β -1,4 ether linkages, which hydrolyze to form xylo-oligomers and eventually xylose. Earlier studies (11) show a biphasic behavior of xylan hydrolysis, a fast breakdown of xylan to xylo-oligomer followed by a slow depolymerization of the residual xylan. The reasons for this biphasic behavior remain elusive. The hydrolysis of low degree of polymerization (DP) xylo-oligomer to xylose is typically fast (11). Reaction (9.1) shows the hydrolysis of a backbone section of xylan by an S_N1 water substitution

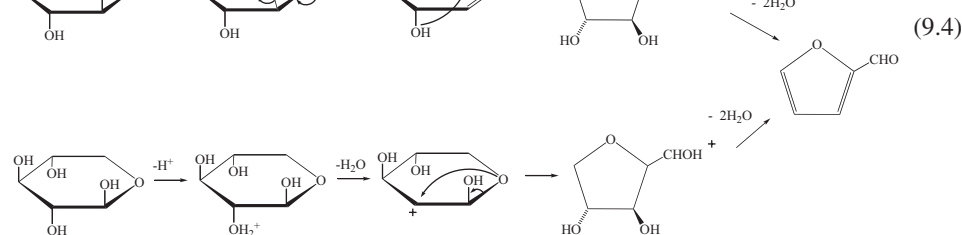
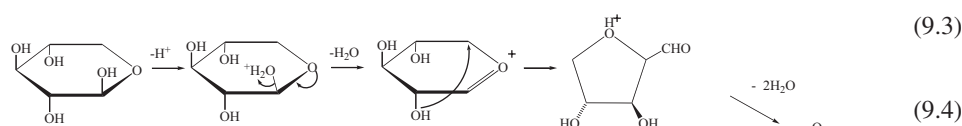
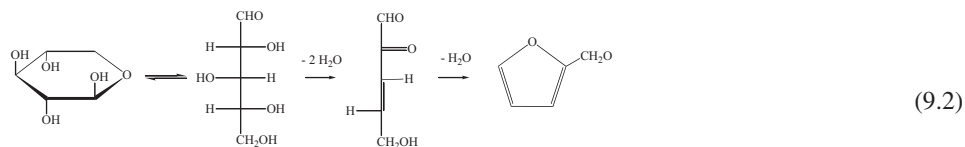
mechanism. The mechanism involves addition of a proton to the oxygen atom of the ether linkage, which leads to the dissociation of the polymer to form a positively charged oxonium ion and a shorter xylan chain. A nearby water molecule quickly interacts with the oxonium ion forming a neutral xylan chain. The extra proton from the water molecule is recycled back into the solution by forming a hydronium ion with the surrounding water molecules. The stability of the oxonium ion suggests fast kinetics for the first half of the hydration reaction. As a result, xylo-oligomers and xylose should be released quickly during the acid pretreatment of xylan and biomass as long as protons are readily available and easily transferred to the ether bond. The rate of protonation of the ether linkage is unknown. Most likely, it will depend on the macroscopic acid concentration and microscopic atomic and molecular environment, which may hinder/promote proton transport to the ether linkage.



The mechanisms of the dehydration or degradation reactions that lead to the destruction of sugar molecules have been less clearly understood. The decomposition of xylose in acidic aqueous solutions has been the focus of a number of studies dating back (12) to the 1930s. Furfural is found to be the main product from the decomposition reactions and this process is used in the industrial production of furfural from oat hulls. The mechanism for furfural formation was initially proposed (13–15) to occur via the open chain form of the sugar structure as is shown in (9.2). Even though xylose molecules have a predominant ring structure in water, the dehydration of the open chain form to furfural drives the equilibrium to the right. Moreover, the protonation of the ring oxygen opens up the ring structure. Degradation proceeds via the elimination of the two water molecules and eventual closure of the open chain to form a furan ring. No intermediate has been detected experimentally to validate this reaction mechanism.

Antal and coworkers (16) proposed another mechanism for xylose degradation to furfural via direct conversion from the six-carbon pyranose ring structure to the five-carbon furan ring structure. The existence of this mechanism was confirmed by recent *ab initio* molecular dynamics simulations and quantum mechanical calculations (17, 18). Two mechanisms, (9.3) and (9.4), were proposed (16) that involve direct rearrangement of the cyclic ring structure after the protonation of the hydroxyl groups and loss of water. Further elimination of the water molecules leads to the formation of furfural. The kinetics of furfural formation from xylose in acid solutions has been measured (19) and the reported activation energy is about 32 kcal mol⁻¹. Direct dehydration or degradation of xylan has not been reported, though degradation of xylo-oligomer, particularly xylo-oligosaccharides, has been reported when certain inorganic salts are added to the solution (9). It is likely that these processes also have relatively high activation energies. Since hydrolysis is facile if protons are readily available for the β -1,4 ether linkages, and dehydration reactions have high activation energies, it is theoretically possible to obtain high yields of xylose during dilute acid pretreatment of xylan and biomass.

However, the low yields of xylose (60–65%) (20–22), and high level of furfural formation (15%), suggest that more needs to be learned about the mechanisms and kinetics of hydrolysis and degradation.



9.2 Computational techniques

In this chapter, quantum mechanical calculations investigating the energetics and kinetics of these reactions are discussed. The hydrolysis of xylobiose and the degradation of xylose were investigated using quantum mechanical molecular dynamics simulations and with static electron structure theory. The molecular dynamics simulations sample a large portion of the potential energy surface of a reacting system and can help identify probable reaction pathways. The sugar degradation reactions were carried out both in the absence and presence of the surrounding water molecules. Static quantum mechanical approaches provide more accurate energies, allowing the determination of reaction energy barriers, which are used to compare the likelihood of competing reaction pathways. These approaches when applied to degradation reactions of xylose provide considerable insight into the likely mechanisms and kinetics. Further application of the static techniques to xylobiose has shed light on the relative reaction rates of hydration and dehydration for xylo-oligomers and xylan.

9.2.1 Molecular dynamics simulations

These calculations are computer simulations of the time evolution of atoms based upon their velocities and potential energies due to interactions with the other atoms in the system. In Car–Parrinello molecular dynamics (CPMD) (23, 24) approach the potential energy due to the interactions of the atoms in the system is determined by conducting quantum mechanical

calculations of the valence and semi-core electrons in each atom. The CPMD code is based on density functional theory and is capable of simulating chemical reaction pathways for systems of up to several thousand atoms. The valence electrons were treated using the density functional developed by Becke (25) and Lee *et al.* (26), BLYP, and were assumed to exist in a pseudopotential exerted by the nuclei and the core electrons. The BLYP functional was shown to be appropriate to describe the liquid water (27, 28). The valence and semi-core electrons were treated with the Troullier-Martin norm-conserving pseudopotential (20). Plane waves were used as the basis functions in these calculations and the simulations were conducted using a time step of 0.125 fs in our calculations. The plane-wave basis set cut-off used was 70 Ry, which was shown to be sufficient for biomolecular simulations in aqueous solution (28). Three-dimensional periodic boundary conditions were applied. During simulations of sugar decomposition reactions in vacuum, the calculations were carried out with a ($12 \times 12 \times 12 \text{ \AA}^3$) unit cell containing one protonated sugar molecule surrounded by sufficient vacuum space to separate the interactions between the neighboring sugar molecules. Protonation was assumed to initiate the sugar degradation reactions. All probable protonation sites of the sugar molecule were investigated including the hydroxyl groups on the sugar ring and the ring oxygen. The simulations were carried out at 500 K. A total of 2 ps simulation time was carried out on a high-speed Linux PC machine. Each step took about 60 seconds of CPU time. The initial atomic coordinates of the xylose molecule were taken from the optimized structure of the CPMD calculations without the proton. The events simulated are qualitative, not quantitative in nature.

Simulations of xylose in water were carried out with each sugar molecule surrounded by 32 water molecules in a unit cell of with a lattice parameter of 11.5 Å. Each sugar molecule had approximately two hydration shells. In addition, one proton was added to the system to mimic the acidic medium. Ab initio MD was carried out at constant temperature of 500 K, which is at the higher end of the pretreatment temperature. Because these calculations sampled a large portion of the conformational space, they were ideally suited for finding low energy structures in sugars and in identifying likely reaction pathways.

9.2.2 Static electronic structure theory

Quantum mechanical calculations were used to obtain energies for reactants, transition states, and products so that reaction energetics and barriers could be obtained. The Gaussian 03 (29) suite of programs were used in our calculations, which were designed to obtain minima in potential energy surfaces corresponding to stable molecular species and saddle points that correspond to transition states. For this study, the hybrid density functional, B3LYP (25, 26), was used to obtain molecular geometries and energies. More accurate energies were obtained for the monosaccharides using the complete basis set (CBS) approach termed CBS-QB3 (30), in which the geometry is optimized using B3LYP/6-311G(d,p) and the energy is extrapolated to the complete basis set limit with MP2. Comparison of results from these techniques to the experimental values for the G2 set of molecules (31) shows that the standard deviation for the B3LYP technique with a split-level basis set (6-31G[d,p]) is about (32) 3 kcal mol⁻¹ while the standard deviation for CBS-QB3 is (30) 1.2 kcal mol⁻¹. The B3LYP technique can underestimate transition states (33–37) by up to 5 kcal mol⁻¹, but the CBS-QB3 approach provides much better estimates of transition state energies (38).

The starting geometries for stable species were selected from the low energy conformers from CPMD calculations and literature results. The optimized geometries of reactants and products had no imaginary vibrational frequencies, whereas transition states had exactly one. Transition states were also confirmed by visual inspection of the motion of the imaginary frequency and by intrinsic reaction coordinate (IRC) calculations (39, 40). Reaction energy barriers, E_a , were determined as the difference in total energies of the transition state and the reactant, including the zero-point energy.

9.3 Xylose degradation reactions in vacuum

Dehydration of xylose was initially investigated by modeling the reactions of protonated β -D-xylose in vacuum (17, 18). The absence of solvent water molecules allows the study of the intrinsic dehydration mechanisms, without the interference of the solvent. Furthermore, the quantum mechanical calculations without solvent molecules are more tractable, particularly for static calculations to obtain the energetics of the reactants, intermediates, and products. In addition, the knowledge of the intrinsic reaction in vacuum is often necessary to model solvated systems.

The simulations started when a proton was added to each of the four hydroxyl groups or the ring oxygen on the xylose ring structure. The subsequent reactions in vacuum were followed using *ab initio* molecular dynamics simulations. The results of these simulations indicate that protonation at O1 and O4 does not lead to any observable xylose degradation during the course of the 2 ps simulations time. Protonation at O5 results in a reversible ring opening and closing reaction. Protonation at O2 results in irreversible dehydration and degradation reaction to form furfural. Protonation at O3 leads to the fragmentation of a xylose molecule to a one-carbon (formic acid) and a four-carbon product. Formic acid has been observed experimentally. Mechanism 3, which results from the protonation at O1, was not observed during *ab initio* MD simulations. Static electronic structure calculations confirmed these simulation results and determined the likely intermediates in the mechanism of furfural formation. Furthermore, these calculations demonstrated that mechanism (9.4) is the most likely route to furfural formation based upon the calculated activation energies of the reaction steps. Figure 9.1 shows an overview of the reactions of these five isomers of a protonated xylose molecule.

Molecular dynamics simulations starting with protonated xylose provide an indication of the reactivity of that protonated isomer. The results of molecular dynamics simulations for protonation (17) at O2 and O3 are shown in Figures 9.2 and 9.3. The colored pictures at the top of these figures show snap shots of the progress of the reaction as a function of simulation time. Lewis structures of the progress of the reactions are shown at the bottom of the figures to provide clarity. As can be seen in Figure 9.2, xylose protonated at O2 dehydrates and rearranges to a furanyl compound within 700 fs. This mechanism is identical to reaction mechanism (9.4) for the formation of furfural proposed by Shafizadeh (41). As shown in Figure 9.3, xylose protonated at O3 dehydrates and breaks the C1–C2 bond within 250 fs. The resulting ether intermediate readily breaks apart to form formic acid and a 1,4-diol. This mechanism accounts for the formation of formic acid that has been reported in the literature (42). The diol that is formed as a by-product in this reaction is likely reactive and may polymerize to form a resin that is commonly found during acid degradation of

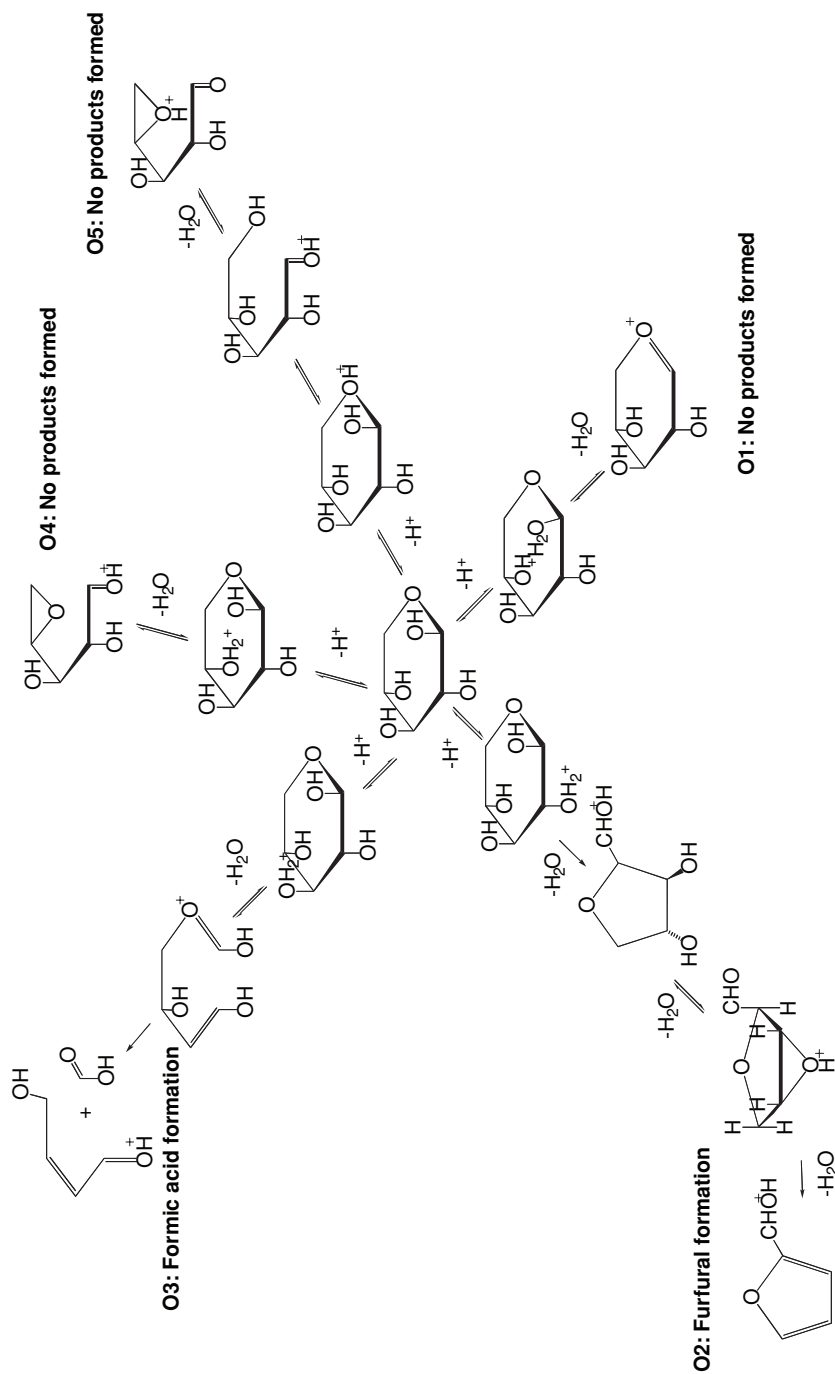


Figure 9.1 Scheme for reactions of protonated xylose as determined by quantum mechanical modeling (CPMD and Gaussian). Protonation at O2 leads to the formation of furfural while protonation at O3 leads to formic acid. Protonations at O1 and O4 lead to dehydration products that readily recombine with water to reform xylose. Protonation at O5 leads to ring opening, which readily re-closes.

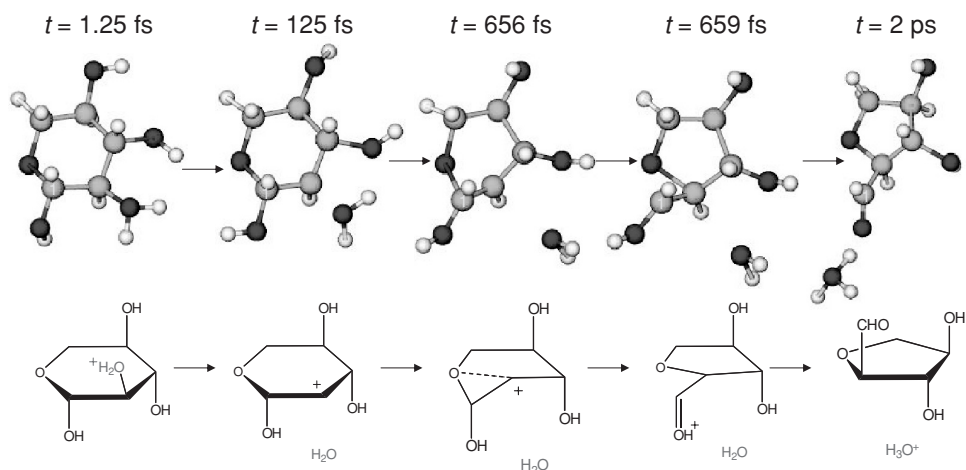


Figure 9.2 Results of CPMD simulation of xylose degradation after protonation of the hydroxyl group on O2. After 125 fs xylose is dehydrated and at approximately 659 fs the remaining carbocation rearranges to form the dehydrated furanyl form of xylose. This product will need to undergo two additional dehydrations to form furfural. (Reproduced in color as Plate 24.)

xylose. Early reports (42) suggest that the formation of resin is accompanied by formic acid formation.

Molecular dynamics simulations of xyloses protonated at the other oxygen atoms were run well beyond 1 ps and showed no net reaction. MD simulations of xylose protonated at O1 did not also lead to the formation of the furanyl product as proposed in reaction

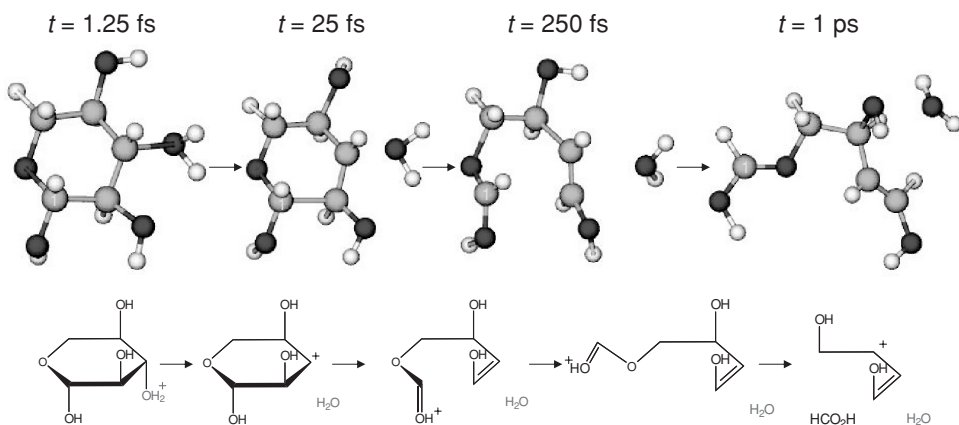


Figure 9.3 Results of CPMD simulation of xylose degradation after protonation of the hydroxyl group on O3. After 25 fs xylose is dehydrated and at approximately 250 fs the C1–C2 bond has broken. In this product the C1–O5 will eventually break to yield formic acid as shown in the Lewis structure at the end. (Reproduced in color as Plate 25.)

Table 9.1 Calculated^a proton affinities (PA) and activation energies for xylose reaction^b

Protonation site	PA (kcal mol ⁻¹)	Activation energies (kcal mol ⁻¹)		
		Step 1	Step 2	Step 3
O1	186.7	4.4		
O2	191.3	16.4	10.7	10.5
O3	188.8	13.6	17.3	
O4	187.2	11.8		
O5	189.5	9.8		

^a Calculated using CBS-QB3 to determine the enthalpy of xylose and the enthalpy of protonated xylose.^b See Figure 9.1

mechanism (9.2). These simulations showed that protonation at O1 leads to dehydration (see Figure 9.1) to form the oxonium, which did not react further. This ion will most likely react quickly with solvent water molecules to reform xylose. Simulations show that protonation and dehydration of xylose at O4 leads to the formation of the species depicted in Figure 9.1 containing a three-atom ring. This species does not react further, but is likely to recombine with water to reform xylose. Of particular interest is the mechanism for the formation of furfural from the open form of xylose, reaction mechanism (9.2). Molecular dynamics simulations show that protonation of O5 leads to the open form of the sugar, which quickly reverts to the cyclic form.

Static electronic structure calculations (CBS-QB3) were used to determine the energetics of the reaction pathways (18) shown in Figure 9.1. The proton affinities, or the gas phase enthalpies to add a proton, provide an indication of the likelihood of the proton to attack the oxygen atoms on xylose. The proton affinities are found to have the following order: O2 > O5 > O3 > O4 > O1 with values shown in Table 9.1. These values suggest that O2 has the highest proton affinity and is more susceptible to protonation, while O1 is least susceptible to protonation. These results show that the proton has a preference for O2, which leads to the formation of furfural. This is consistent with the experimental observation that acid treatment of xylose leads primarily to furfural. Static electronic structure calculations of the activation energies for the steps shown in Figure 9.1 are also consistent with the observations from the MD simulations. Table 9.1 also lists the calculated activation energies for the steps shown in this figure. Protonation at O1 readily leads to the oxonium ion as confirmed by the low barrier. However, the subsequent reaction to form the furanyl compound shown in Reaction (9.3) has a high-calculated barrier (32.0 kcal mol⁻¹). Protonation at O2 has a low barrier for the first step and all subsequent steps. This is consistent with experimental studies that have failed to observe any intermediates. Protonation at O3 has low barriers to form formic acid, consistent with the experimental observation of this product. No reactions were found from the protonation at O4. Protonation at O5 readily leads to ring opening, but subsequent reactions that lead to the formation of furfural have high barriers (18) (27–29 kcal mol⁻¹). Thus, the original mechanism proposed for furfural formation, Reaction mechanism (9.2), appears to be less likely than Reaction mechanism (9.4), which has barriers of 10–17 kcal mol⁻¹. These results also suggest that protonation at O2 leads to furfural formation and protonation at O3 leads to formic acid,

both of which have been observed experimentally. NMR measurements (18) also confirm these two products and experiments with ^{13}C labeled are consistent with the proposed mechanisms.

9.4 Effects of solvent water molecules

Similar to the simulations carried out in vacuum (14), *ab initio* MD simulations of xylose degradation in water started when a protonated β -D-xylose was positioned in a unit cell surrounded by 32 water molecules (43). During the course of our entire simulation (~ 5 ps), the initiation proton attached to the $-\text{OH}$ groups on the sugar ring was observed to be transferred back to the surrounding water molecules. This transfer is rapid and occurs in less than 100 fs for all of the hydroxyl groups on the xylose ring. Once the proton was transferred to the neighboring water molecule, it is quickly transferred to other water molecules and away from the sugar molecule. This result shows that protonation is probably the rate-limiting step in sugar degradation under acidic media because our earlier simulations in vacuum demonstrate that protonated β -D-xylose molecules decompose rapidly. Figure 9.4 shows the proton transfer from xylose C2-OH to the solvent water molecules from our simulations. The simulations started with a protonated xylose molecule surrounded by 32 water molecules. After 34 fs, a neighboring water molecule forms a bond with the protonated hydroxyl group ($\text{C2-OH}_2^+-\text{OH}_2$) as shown in the figure. After 62 fs, the proton from the C2-OH transfers to a water molecule, forming an H_3O^+ . After the proton was transferred from the xylose hydroxyl group to the water molecule, it was quickly transferred to other water molecules and away from the xylose molecule due to the strong hydrogen bonding interactions between the water molecules and the high proton mobility. It appears that protonation is a slow rate-limiting step.

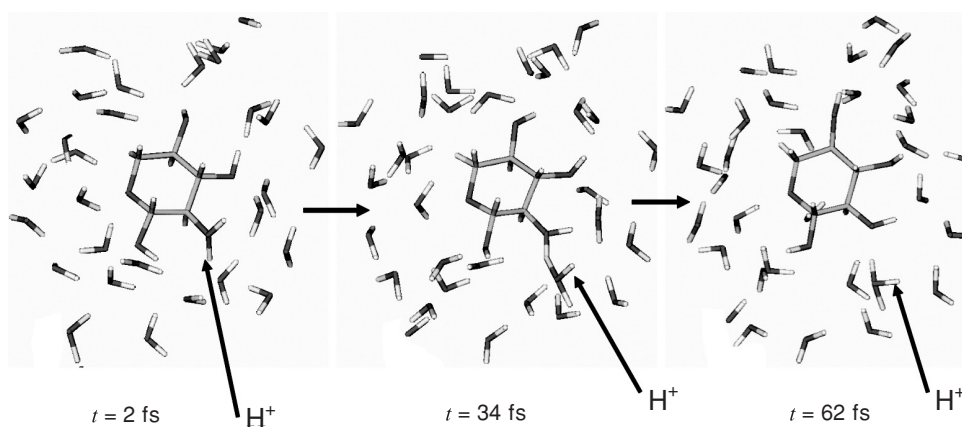


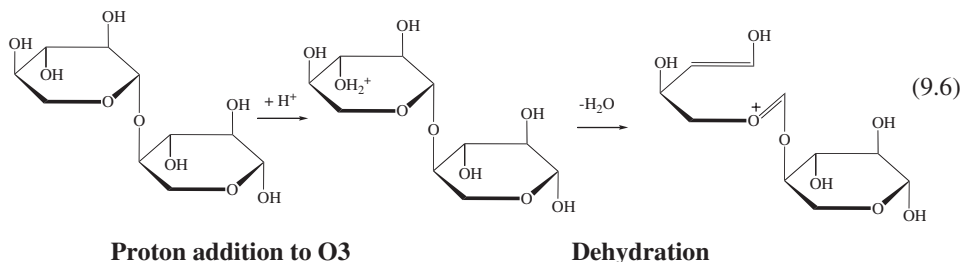
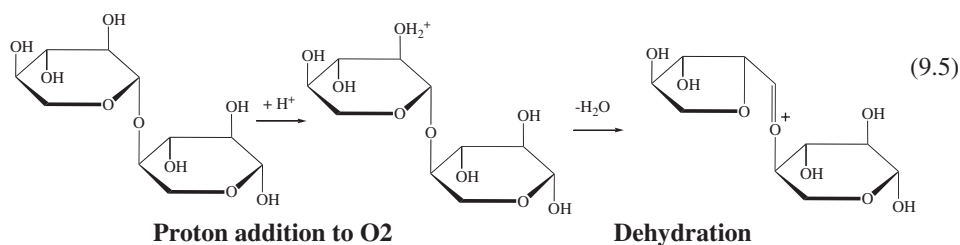
Figure 9.4 Snapshots of the MD simulations showing the rapid proton transfer from a xylose molecule to water. (Reproduced in color as Plate 26.)

In order to test the notion that protonation is the rate-limiting step, reaction barriers have been estimated using the hybrid density functional B3LYP (18). A single water molecule has a lower proton affinity (44), $PA = 165 \text{ kcal mol}^{-1}$, than xylose (Table 9.1 $PA = 186.7\text{--}191.3 \text{ kcal mol}^{-1}$), but the MD simulations appear to indicate that a proton will be transferred from xylose to a neighboring water cluster. CBS-QB3 calculations (18) show that for water clusters, the proton affinity increases with cluster size due to the increased stability of the hydronium ion. These calculations show that for a four-molecule cluster, the proton affinity has increased to $220.2 \text{ kcal mol}^{-1}$. This and the MD simulations suggest that bulk water will have a higher proton affinity than xylose. The high proton affinity of bulk water relative to xylose must then add energy to the intrinsic energy barrier for the dehydration of xylose. This could explain the discrepancy between the energy barriers calculated in vacuum shown in Table 9.1 ($\sim 16 \text{ kcal mol}^{-1}$ for the formation of furfural) and the experimentally observed barriers in solution, about 32 kcal mol^{-1} . This was investigated further with static electronic structure calculations using B3LYP. Energy barriers for the initial step of xylose conversion to furfural were calculated in the presence of water clusters. With water clusters present, the barrier for xylose dehydration increased (18) to around 30 kcal mol^{-1} , which is consistent with experimental values. This suggests that the experimentally measured barrier for xylose dehydration contains an energy contribution due to transferring the proton from the solvent to the substrate.

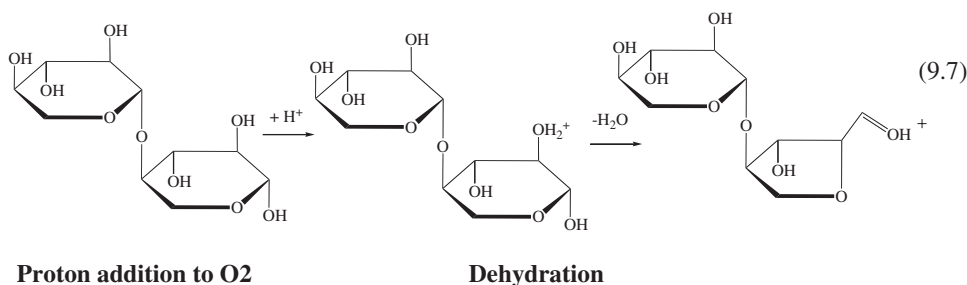
9.5 Xylobiose calculations

As a model system for hydrolysis and dehydration of xylo-oligomers and xylan, static quantum mechanical calculations were conducted on the decomposition of xylobiose. As with xylose, acid-catalyzed reactions were studied by computing the energy barriers for protonated xylobiose in vacuum. Because of the large size of this molecule, highly accurate CBS calculations could not be conducted. Instead DFT [B3LYP/6-31(d,p)] calculations were conducted. This level of theory could underestimate (33–37) reaction barriers by 5 kcal mol^{-1} , but is sufficient for a semi-quantitative comparison of the barriers for different reaction pathways. Initial molecular structures of neutral xylobiose, the protonated reactants, and the transition states used the corresponding structures from the study of xylose (17, 18) discussed above. Likewise, analogous reaction mechanisms were studied for xylobiose. For example, Reactions (9.5) and (9.6) show the first steps in the decomposition of xylobiose protonated O2 and O3 on the non-reducing end to form a furanyl ring or the precursor to formic acid. Reaction (9.5) exhibits a xylobiose dehydration mechanism similar to a monomer xylose (Figure 9.1). The degradation of the xylobiose is initiated when O2 on the non-reducing end is protonated. The protonated hydroxyl group (i.e., H_2O) leaves the sugar ring forming a carbocation, which reorganizes to form an uncharged five-member ring leaving the positive charge outside the ring structure. A water molecule will then likely hydrolyze the ether linkage to break the dimer into a furan ring and an intact xylose molecule. The further dehydration of the furan ring structure leads to the formation of furfural. Likewise, the reaction could also initiate by protonation of O3 on the non-reducing end of xylobiose in a similar mechanism to the xylose monomer. However, here the

product with the open structure will dissociate into a xylose ester and a four-carbon cation product.



Protonation at O2 on the reducing end of xylobiose results in the reaction shown in (9.7), while protonation at O3 on the reducing end produces an unstable structure, which without a barrier, transfers its proton to O5 on the non-reducing end. The product from Reaction (9.7) might then dehydrate through reactions similar to those shown in Figure 9.1 to form a furfural molecule and a xylose molecule.



In addition to the dehydration reaction shown in Reactions (9.5)–(9.7), the hydrolysis reaction to form two xylose molecules was considered. For this reaction, a proton is added to the ether linkage, which decomposes to a xylose molecule and an oxonium ion as shown in Reaction (9.8). In aqueous solution, the oxonium cation will be quickly hydrolyzed to form xylose as was discussed earlier. A comparison of the barrier for this process to the barriers for dehydration reactions shown, (9.5)–(9.7), was used to determine which reaction pathways are most likely. Importantly, no reaction barrier could be found for the hydrolysis reaction (9.8). If a proton was added to the ether linkage, the molecule decomposed without a barrier into xylose and the oxonium ion shown in Reaction (9.8). These calculations were conducted

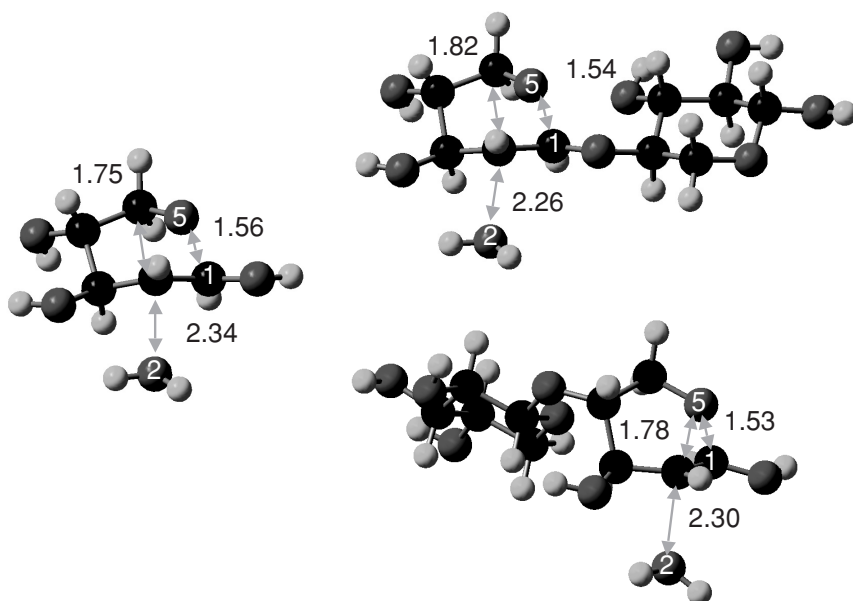


Figure 9.5 Transition states for dehydration reaction of xylose (left) and xylobiose (right-top and side view) resulting from protonation at O2. (Reproduced in color as Plate 27.)

several times and always led to the same result. On the other hand, the reaction barriers for the dehydration reactions were found to be significantly larger and were consistent with the barriers for the dehydration reactions of xylose monomer. Using B3LYP/6-311G(d,p), barriers of 17.8 and 19.8 kcal mol⁻¹ were obtained for Reactions (9.5) and (9.6). These are close to the barriers calculated for similar reactions for neat xylose (Table 9.1). Figures 9.5 and 9.6 compare the calculated molecular geometries for the transition states for dehydration of xylose and xylobiose protonated at O2 and O3. Notice that the bond lengths at the reacting centers are similar for xylose and xylobiose. The barrier calculated for Reaction

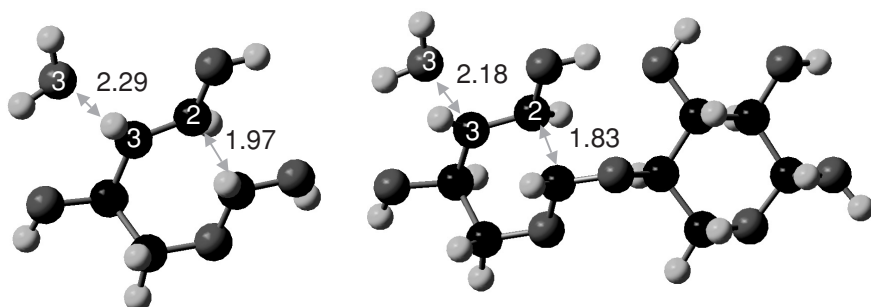
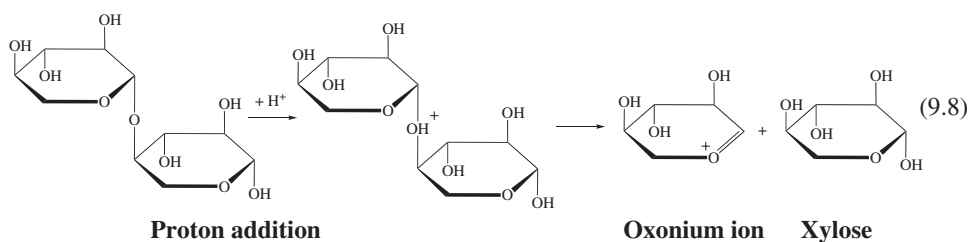


Figure 9.6 Transition states for dehydration reaction of xylose (left) and xylobiose (left) after protonation at O3. (Reproduced in color as Plate 28.)

(9.7) is lower ($4.7 \text{ kcal mol}^{-1}$), but is still significantly higher than the barrier for Reaction (9.8).



The absence of a barrier for Reaction (9.8) suggests that this process is kinetically favored over the dehydration reactions (9.5)–(9.7). This is not surprising, because, as mentioned above, Reaction (9.8) results in the formation of a relatively stable oxonium. Figure 9.7 shows a comparison of calculated reaction barriers for xylobiose protonated at the O2 and O3 sites and protonated on the linker oxygen atom. These calculations were conducted in vacuum, and as we have shown in our calculations of xylose, the barrier in aqueous solution is likely to be higher due to the endothermicity of the proton transferring from the solvent water molecules to the sugar oxygen atoms. However, since ethers have higher proton affinities than alcohols (44), the increase in the barrier due to solvation should be less for hydrolysis reactions (9.8), than for dehydration reactions, (9.5)–(9.7). Our calculations suggest that the barrier for hydrolysis of xylobiose should be significantly lower than the barriers for dehydration reactions and that the kinetics of hydrolysis, consequently, should dominate. Because the ether linkage is identical in other xylo-oligomers, this result further suggests that for all xylo-oligomers, hydrolysis should dominate. Loss of xylose due to dehydration reactions should only result from the dehydration of xylose itself, not from dehydration reactions of the xylo-oligomers when there is only acid in the solution. The accelerate destruction of xylose and xylotriose molecules with the addition of

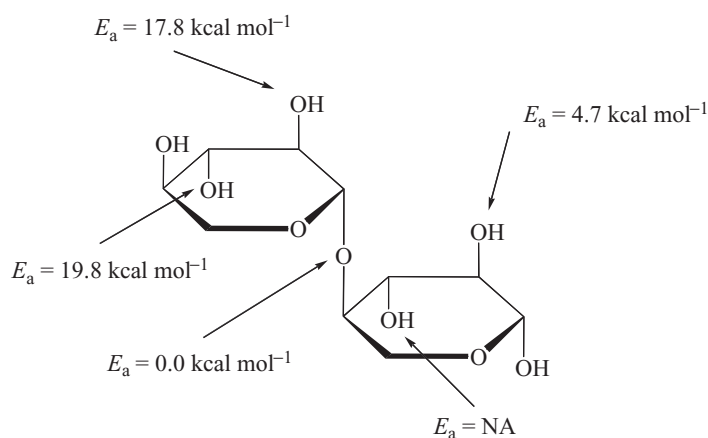


Figure 9.7 Calculated reaction barrier energies for xylobiose protonated at the ether linkage and the O2 and O3 atoms. No barrier was obtained for protonation at O3 on the right xylose residue because the protonated species was unstable. It spontaneously transfers a proton to O5 on the left residue.

inorganic salts (9) is likely due to a catalytic function, which lowers the barrier of dehydration reactions.

9.6 Experimental investigation of hydrolysis

In order to investigate the relative rates of hydrolysis and dehydration, experiments were conducted on xylose, xylan, and xylobiose using a small, glass reactor that was heated with microwave energy [CEM-Discover]. In these experiments, products were measured using high pressure liquid chromatography (HPLC). The microwave reactor system consists of an 8 mL glass tube enclosed inside the cavity of a microwave heating unit. Experiments were conducted with 2 mL of solution containing the substrate in an aqueous solution of sulfuric acid (1.2% by weight). The tube was fitted with a Teflon-coated cap and a pressure sensor, which are designed to contain pressures up to 250 psi. The tubes contained a Teflon-coated stir bar and the temperature was measured with an optical pyrometer. Batch experiments were conducted in which the samples were heated to a fixed temperature.

Hydrolyzates were subjected to chemical analysis using HPLC to determine the concentrations of xylose, xylobiose, and xylose degradation products present in the reaction solutions. The solutions were analyzed using an Agilent 1100 series HPLC with an HPX-87H column and a precolumn (Bio-Rad Laboratories) operated at 65°C. The eluant was 0.01N H₂SO₄ flowing at 0.6 mL/min. Samples and standards were injected (10 µL) onto the column after filtering through a 0.45 µm nylon membrane filter (Pall, Acrodisc Syringe filter). Solute concentrations were measured with an Agilent 1100 refractive index detector controlled to 45°C and a diode array detector. The detectors were calibrated with a set of four standards for all solutes except xylobiose, which had a single-point calibration. The HPLC was controlled and data was analyzed using Agilent Chemstation software (rev A.09.03).

Before the microwave heating system was used for kinetic measurements of xylobiose and xylan decomposition, an accurate temperature in the reactor was obtained. The provided optical pyrometer measures the infrared light emitted from the reactor and could provide an inaccurate temperature if the walls of the reactor were cooler than the solution. A more accurate technique for measuring the temperature would be to use a chemical reaction with known activation energy (chemical thermometer). In this study, we used the thermal decomposition of xylose in acid solutions as our chemical thermometer. We measured the decomposition of xylose at a fixed nominal temperature and compared the measured rate constant to the values reported in the literature to extract an effective temperature. The relationship between the rate constant for the decomposition of xylose and the temperature has been reported (19) to be

$$k = 0.0453a\delta\gamma_x C_{11} e^{-35.70\left(\frac{4731.7-T}{T}\right)} \quad (\text{Equation I})$$

where a is the ratio of the density of xylose solution to that of the solution without xylose ($a = 1$), δ is the specific gravity of water at a given temperature relative to the specific gravity at 30°C, γ_x is a correlation constant that was empirically determined ($\gamma_x = 0.95$), C_{11} is the acid concentration, and T is the absolute temperature.

Table 9.2 Measured rate constants and temperature of microwave reactor^a

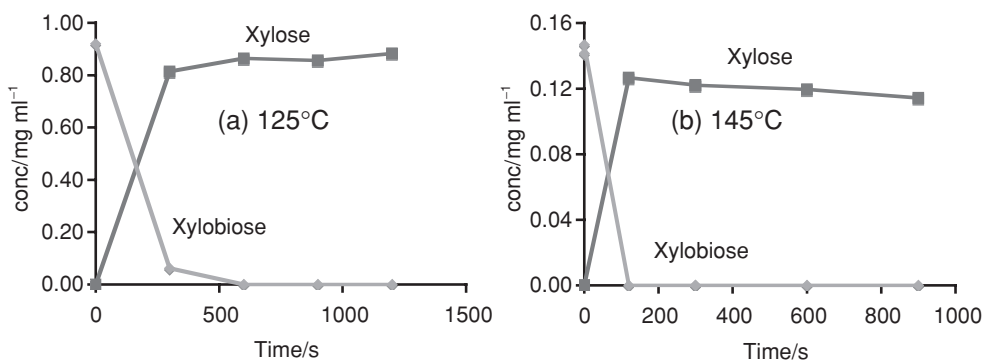
Nominal temperature (°C)	Rate constant (s ⁻¹)	Effective temperature (°C)
125	3.7×10^{-5}	117
135	2.5×10^{-5}	134
145	1.0×10^{-4}	149
155	3.7×10^{-4}	163
165	6.7×10^{-4}	170
175	2.9×10^{-3}	188

^a Determined from Equation I.

These data were then used in Equation I to determine the actual reactor temperatures, which are shown in Table 9.2. As Table 9.2 shows, the effective temperature is typically slightly higher than the nominal temperature measured by the optical pyrometer. Since the microwaves heat the solution directly, it is reasonable that the glass reactor tube would have a lower temperature than the solution. The optical pyrometer measures the temperature on the surface of the glass and it is not surprising that the nominal temperature is lower than the effective temperature in the solution.

9.7 The hydrolysis of xylobiose

Experiments were conducted with xylobiose in the microwave reactor and results are shown in Figure 9.8 for the decomposition of xylobiose and the formation of xylose at nominal temperatures of 125 and 145°C. As can be seen from these plots, the xylobiose is quickly converted to xylose at these temperatures, i.e., within 600 seconds for 125°C and within 120 seconds for 145°C. Furfural formation was less than that observed for decomposition of pure xylose, suggesting that the furfural arose from the decomposition of xylose and not xylobiose. The mass balances in these experiments were poor (80–85%), due probably to the poor calibration of xylobiose. Because of the expense of this compound, we were only able to use single-point calibrations. However, as Figure 9.9 shows, the chromatographs were clean and did not indicate the formation of other products. These results seem to confirm the

**Figure 9.8** Decomposition of xylobiose and formation of xylose at (a) 125°C and (b) 145°C.

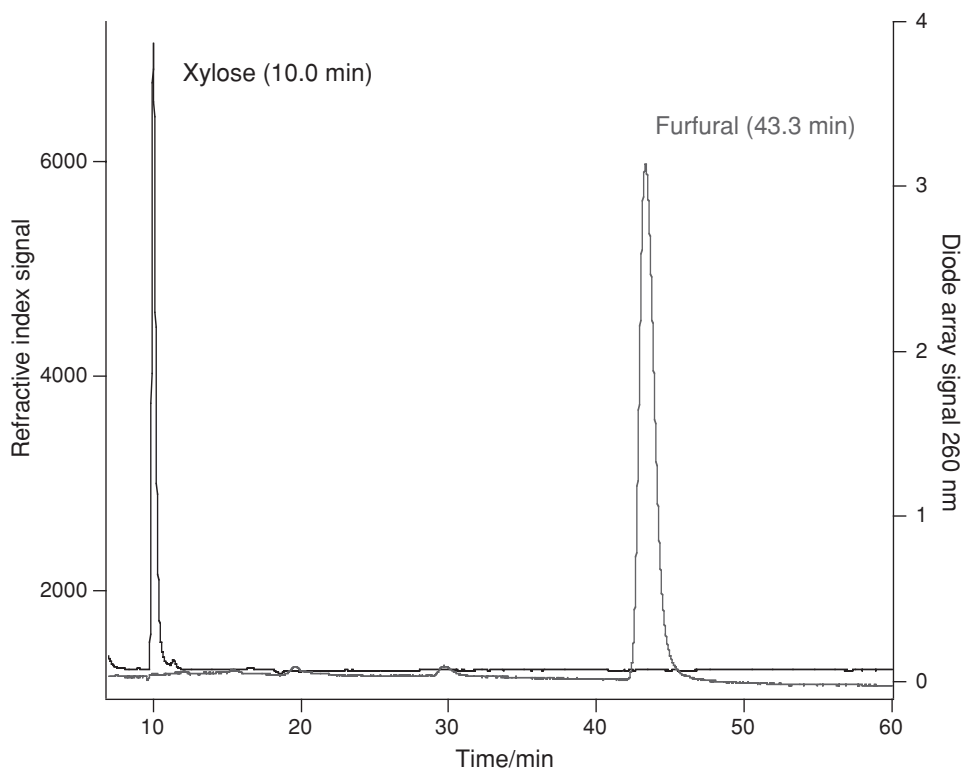


Figure 9.9 HPLC chromatograph of the products from the decomposition of xylobiose at 125°C after 20 minutes. The black trace is from the refractive index detector and the grey trace is from the diode array detector.

results from our calculations that xylobiose will preferentially undergo hydrolysis relative to dehydration. We conclude that low DP xylo-oligomers will preferentially undergo hydrolysis to form smaller oligomers and xylose as opposed to dehydration to form furans in solution containing only catalytic acids.

9.8 The hydrolysis of xylan

In order to investigate the formation and destruction of xylo-oligomers further, experiments were conducted using xylan. Both Birchwood xylan and Beechwood xylan were used for these studies. These materials were supplied as fine powders from Sigma and used as received. Because this biopolymer is not soluble, stock solutions of the samples in acid could not be prepared. Instead, approximately 20 mg of sample was added to 2 mL of a 1.2% w/w aqueous solution of sulfuric acid and the measured products were normalized to the actual amount of xylan used. In the absence of heating, HPLC measurements of the liquid portion of these slurries detected no xylose or furfural, while these species were observed when the samples were heated in the microwave reactor. Figure 9.10 shows plots of xylose and furfural

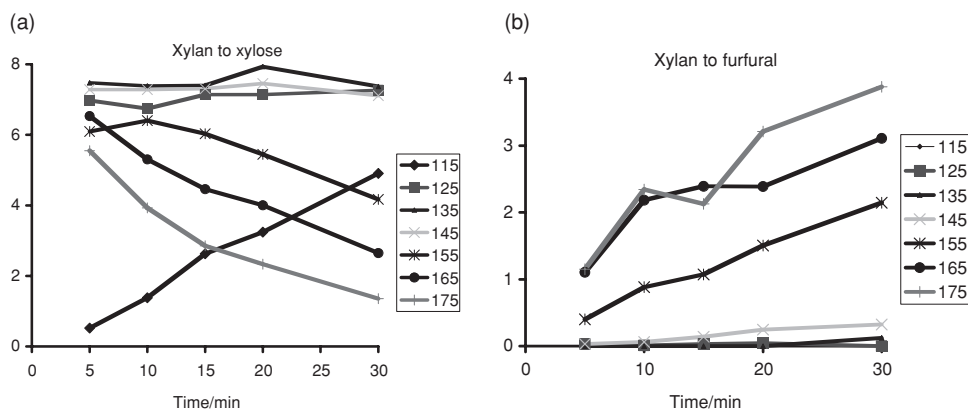


Figure 9.10 Products from acid decomposition of Beechwood xylan slurries. Plots showing the formation of (a) xylose and (b) furfural as a function of residence time for several nominal temperatures.

concentrations (normalized to 10 mg/mL of added xylan) as a function of residence time of the xylan slurry in the reactor. Figure 9.10a shows the plot of xylose formation at 115°C, where it is apparent that increasing the residence time results in increased formation of xylose. This result suggests that, at this temperature, xylan has not been fully converted to xylose. At 125, 135, and 145°C, the amount of xylose formed remains constant at all measured residence times, which suggests that the xylose residues in the xylan have been completely released. We note that approximately 73% of the mass in the xylan has been converted to xylose in these experiments. This value is close to the known concentration of xylose in this xylan [Johnson, D.K. (2006) unpublished work] (69%). At higher temperatures, the concentration of xylose decreases with increasing residence time. This observation suggests that at these temperatures, the xylose is being dehydrated to form furfural, which is also confirmed by the furfural plot in Figure 9.10b. Note that at temperatures below 145°C, independent measurements with pure xylose in acid show little formation of furfural, whereas at these temperatures large amounts of xylose are formed from xylan. Xylobiose was only observed at 115°C and only at a normalized level of 0.002 mg mL⁻¹. These observations again show that hydrolysis reactions of xylo-oligomers and xylan are much faster than dehydration reactions of xylose.

9.9 Corn stover

As a final investigation of xylo-oligomer formation and destruction, samples of powdered corn stover were hydrolyzed in the microwave reactor vessels. These samples were Pioneer 34M95 maize harvested in 2002, which was ground to about 1 mm particle size. Earlier studies have shown that these corn stover samples contain approximately 20% by weight of xylan (10). Approximately, 85 mg of corn stover was used in 20 mL of 1.2% sulfuric acid so that the amount of xylan matched that in the experiments discussed above. Figure 9.11 shows the amount of xylose, xylobiose, and furfural measured as a function of reaction temperature for residence times of 10 and 30 minutes. These experiments were conducted to compare the results using biomass with those using pure xylobiose and xylan. Note that these residence

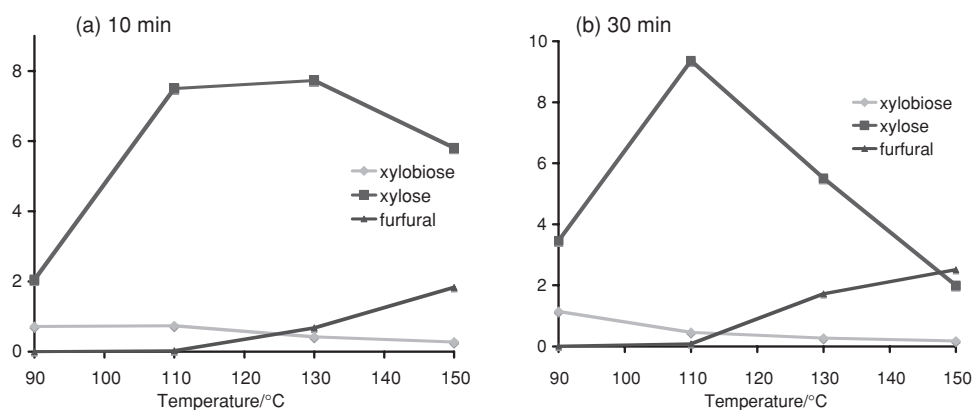


Figure 9.11 Formation of xylose, xylobiose, and furfural as a function of temperature from the acid treatment of corn stover in a microwave reactor. Residence time was (a) 10 minutes and (b) 30 minutes.

times are longer than what is realistic for a commercial scale reactor (2 min). We observed that xylose formation was again maximized at low temperature (135°C), whereas higher temperatures result in the dehydration of xylose and the formation of furfural. Xylobiose formation was minor. These results are completely consistent with the experimental results observed for xylobiose and xylan, and with the theoretical results for xylobiose.

However, it should be noted that the results observed here are different from what has been measured in larger scale pretreatment studies. In those examples, higher biomass loadings have often been used and in some cases no stirring was possible. In these larger scale studies, optimal yields of xylose were found at higher temperatures (>150°C). It seems likely that higher temperatures were required because the mass transport was more restricted in those studies. In this study, a low loading, small particle size, and high agitation of the solution helped reduce mass transport limitations. In our experiments, there was more effective delivery of the catalyst (protons) to the substrate.

9.10 Conclusions

The results of this study suggest that the rates of direct dehydration of xylo-oligomers to form unwanted products during pretreatment is likely to be insignificant compared to hydrolysis to form smaller oligomers and xylose. This was shown both by experiment and theory. Quantum mechanical calculations showed that the barrier for hydrolysis is much smaller than the barriers for possible dehydration reactions. Experimental measurements showed that at low temperatures, the glycosidic linkages in xylo-oligomers can be completely hydrolyzed with little formation of dehydration products.

The results of our experiments with the microwave reactor also showed that the temperature required for complete conversion of xylan and xylo-oligomers to xylose is much lower than required for optimal xylose formation in larger scale studies with corn stover, although the residence times used in our study were larger (5 min and longer compared to 2 min). In this study, optimized yields of xylose were obtained at about 135°C, whereas temperatures greater than 180°C are typically used in pilot scale studies. For our experiments at 135°C,

little conversion to furfural was observed, whereas in pilot scale studies, high conversion at 180°C is accompanied by significant loss of xylose. The most likely explanation for these different observations is the better mass transport available in our microwave reactor-based study. For example, the particle sizes were smaller, the slurries had more water, and the samples were vigorously stirred. Thus, it is likely that the measurements in this study are more likely to result from the intrinsic kinetics of xylan hydrolysis.

An important consequence of these observations is that higher yields of xylose and lower losses to dehydration can be achieved if mass transport can be improved. It appears that further investigations into this area are warranted.

9.11 Future studies

Since mass transport appears to be an important limitation in the hydrolysis of xylan and xylo-oligomers, studies should be conducted on different approaches to increase the access of the acid to the hemicellulose in biomass. It is unlikely that the mass transport available in the experiments in this study could be obtained in a practical pretreatment reactor; however, there might be ways to add mechanical energy during pretreatment that could offset the reaction severity and lead to higher xylose yields and less xylose loss. Experiments and calculations investigating the effects due to increased particle size, concentration and agitation should be conducted. These studies should include molecular modeling to understand proton transfer and sugar recombination reactions and computational fluid dynamics to study fluid flow and mass transfer in relevant particle sizes. The goal of future work should be to develop quantitative kinetic parameters that can be used to design improved pretreatment reactors and pretreatment economics.

Acknowledgment

This work was supported by the US DOE Office of the Biomass Program.

References

1. *Breaking the Biological Barriers to Cellulosic Ethanol*. Department of Energy, DOE/SC-0095, June 2006.
2. *Biofuels in the European Union: A Vision for 2030 and Beyond*, 2006.
3. Luo, C., Brink, D.L. & Blanch, H.W. (2002) Identification of potential fermentation inhibitors in conversion of hybrid poplar hydrolyzate to ethanol. *Biomass and Bioenergy*, **22**, 125–138.
4. Cantarella, M., Cantarella, L., Gallifuoco, A., Spera, A. & Alfani, F. (2004) Effect of inhibitors released during steam-explosion treatment of poplar wood on subsequent enzymatic hydrolysis and SSF. *Biotechnology Progress*, **20**, 200–206.
5. Torget, R.W., Kim, J.S. & Lee, Y.Y. (2000) Fundamental aspects of dilute acid hydrolysis/fractionation kinetics of hardwood carbohydrates. 1. Cellulose hydrolysis. *Industrial and Engineering Chemistry Research*, **39**, 2817–2825.
6. Kim, J.S., Lee, Y.Y. & Torget, R.W. (2001) Cellulose hydrolysis under extremely low sulfuric acid and high-temperature conditions. *Applied Biochemistry and Biotechnology*, **91–93**, 331–340.

7. Xiang, Q., Kim, J.S. & Lee, Y.Y. (2003) A comprehensive kinetic model for dilute-acid hydrolysis of cellulose. *Applied Biochemistry and Biotechnology*, **105–108**, 337–352.
8. Bergeron, P.W., Wright, J.D. & Wyman, C.E. (1989) Dilute acid hydrolysis of biomass for ethanol production. *Energy from Biomass and Wastes*, 1277–1296.
9. Liu, C.G. & Wyman, C.E. (2006) The enhancement of xylose monomer and xylotriose degradation by inorganic salts in aqueous solutions at 180°C. *Carbohydrate Research*, **341**, 2550–2556.
10. Wiselogle, A., Tyson, S. & Johnson, D.K. (1996) Biomass feedstock resources and composition. In: *Handbook on Bioethanol: Production and Utilization* (ed. C.E. Wyman). Taylor and Francis, Washington, DC.
11. Torget, R., Werdene, P., Himmel, M. & Grohmann, K. (1990) Dilute acid pretreatment of short rotation woody and herbaceous crops. *Applied Biochemistry and Biotechnology*, **24–25**, 115–126.
12. Hurd, C.D. & Isenhour, L.L. (1932) Pentose reactions. I. Furfural formation. *Journal of American Chemical Society*, **54**, 317.
13. Harris, D.W. & Feather, M.S. (1973) Evidence for a C-2 \rightarrow C1 intramolecular hydrogen transfer during the acid-catalyzed isomerization of D-glucose to D-fructose. *Carbohydrate Research*, **30**, 359–365.
14. Feather, M.S., Harris, D.W. & Nichols, S.B. (1972) Routes of conversion of D-xylose, hexuronic acids, and L-ascorbic acid to 2-furaldehyde. *Journal of Organic Chemistry*, **37**, 1606–1608.
15. Feather, M.S. (1970) The conversion of D-xylose and D-glucuronic acid to 2-furaldehyde. *Tetrahedron Letters*, **48**, 4143–4145.
16. Antal, M.J., Leesomboon, T., Mok, W.S. & Richards, G.N. (1991) Mechanism of formation of 2-furaldehyde from D-xylose. *Carbohydrate Research*, **217**, 71–85.
17. Qian, X., Nimlos, M.R., Johnson, D.K. & Himmel, M.E. (2005) Acidic sugar degradation pathways: An ab initio molecular dynamics study. *Applied Biochemistry and Biotechnology*, **121–124**, 989–997.
18. Nimlos, M.R., Qian, X., Davis, M., Himmel, M.E. & Johnson, D.K. (2006) Energetics of xylose decomposition as determined using quantum mechanics modeling. *Journal of Physical Chemistry A*, **110**, 11824–11838.
19. Root, D.F., Saeman, J.F. & Harris, J.F. (1959) Kinetics of the acid-catalyzed conversion of xylose to furfural. *Forest Products Journal*, 158.
20. Heitz, M., Capek-Menard, E., Korberle, P.G., Grange, J., Chornet, E., Overend, R.P., Taylor, J.D. & Yu, E. (1991) Fractionation of *Populus tremuloides* at the pilot plant scale: Optimization of steam pretreatment conditions using the STAKE II technology. *Bioresource Technology*, **25**, 23–32.
21. Abatzoglou, N., Chronet, E., Belkacemi, K. & Overend, R.P. (1992) Phenomenological kinetics of complex systems: The development of a generalized severity parameter and its application to lignocellulosics fraction. *Chemical Engineering Science*, **47**, 1109–1122.
22. Saddler, J.N., Ramo, L.P. & Breuil, C. (1993) Steam pretreatment of lignocellulosic residues. *Bioconversion of Forest and Agricultural Plant Residues*. pp. 73–92. CAB International, Oxford.
23. CPMD3.7. (2003) Copyrighted jointly by IBM Corp and by Max-Planck Institute, Stuttgart.
24. Car, R. & Parrinello, M. (1985) Unified approach for molecular-dynamics and density-functional theory. *Physical Review Letters*, **55**, 2471–2474.
25. Becke, A.D. (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, **38**, 3098–3100.
26. Lee, C.T., Yang, W.T. & Parr, R.G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Physical Review B*, **37**, 785–789.
27. Sprik, M., Hutter, J. & Parrinello, M.J. (1996) Ab initio molecular dynamics simulation of liquid water: Comparison of three gradient-corrected density functionals. *Chemical Physics*, **105**, 1142–1152.
28. Molteni, C. & Parrinello, M. (1998) Glucose in aqueous solution by first principles molecular dynamics. *Journal of the American Chemical Society*, **120**, 2168–2171.

29. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, J.J.A. Vreven, T., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Gonzalez, C. & Pople, J.A. (2001) *Gaussian 03*. Gaussian, Inc, Wallingford, CT.
30. Montgomery, J.A., Frisch, M.J., Ochterski, J.W. & Petersson, G.A. (1999) A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *Journal of Chemical Physics*, **110**, 2822–2827.
31. Curtiss, L.A., Raghavachari, K., Trucks, G.W. & Pople, J.A. (1991) Gaussian-2 theory for molecular-energies of 1st-row and 2nd-row compounds. *Journal of Chemical Physics*, **94**, 7221–7230.
32. Foresman, J.B. & Frisch, A. (1995) *Exploring Chemistry with Electronic Structure Methods*. Gaussian, Inc, Pittsburgh, PA.
33. Oie, T., Topol, I.A. & Burt, S.K. (1995) Ab-initio and density-functional studies on internal-rotation and corresponding transition-states in conjugated molecules. *Journal of Physical Chemistry*, **99**, 905–915.
34. Hodgson, D., Zhang, H.Y., Nimlos, M.R. & McKinnon, J.T. (2001) Quantum chemical and RRKM investigation of the elementary channels of the reaction $C_6H_6 + O(P-3)$. *Journal of Physical Chemistry A*, **105**, 4316–4327.
35. Basch, H. & Hoz, S. (1997) Ab initio study of hydrogen abstraction reactions. *Journal of Physical Chemistry A*, **101**, 4416–4431.
36. Bach, R.D., Glukhovtsev, M.N., Gonzalez, C., Marquez, M., Estevez, C.M., Baboul, A.G. & Schlegel, H.B. (1997) Nature of the transition structure for alkene epoxidation by peroxyformic acid, dioxirane, and dimethyldioxirane: A comparison of B3LYP density functional theory with higher computational levels. *Journal of Physical Chemistry A*, **101**, 6092–6100.
37. Dobbs, K.D. & Dixon, D.A. (1994) Ab-initio prediction of the activation-energy for the abstraction of a hydrogen-atom from methane by chlorine atom. *Journal of Physical Chemistry*, **98**, 12584–12589.
38. Nimlos, M.R., Blanksby, S.J., Ellison, G.B. & Evans, R.J. (2003) Enhancement of 1,2-dehydration of alcohols by alkali cations and protons: A model for dehydration of carbohydrates. *Journal of Analytical and Applied Pyrolysis*, **66**, 3–27.
39. Gonzalez, C. & Schlegel, H.B. (1990) Reaction-path following in mass-weighted internal coordinates. *Journal of Physical Chemistry*, **94**, 5523–5527.
40. Gonzalez, C. & Schlegel, H.B. (1989) An improved algorithm for reaction-path following. *Journal of Chemical Physics*, **90**, 2154–2161.
41. Shafizadeh, F., McGinnis, G.D. & Philpot, C.W. (1972) Thermal degradation of xylan and related model compounds. *Carbohydrate Research*, **25**, 23–33.
42. Dunlop, A.P. (1948) Furfural formation and behavior. *Industrial and Engineering Chemistry*, **40**, 204–209.
43. Qian, X.H., Nimlos, M.R., Davis, M., Johnson, D.K. & Himmel, M.E. (2005) Ab initio molecular dynamics simulations of beta-D-glucose and beta-D-xylose degradation mechanisms in acidic aqueous solution. *Carbohydrate Research*, **340**, 2319–2327.
44. Hunter, E.P.L. & Lias, S.G. (1998) Evaluated gas phase basicities and proton affinities of molecules: An update. *Journal of Physical and Chemical Reference Data*, **27**, 413–656.

Chapter 10

Enzymatic Depolymerization of Plant Cell Wall Hemicelluloses

*Stephen R. Decker, Matti Siika-aho, and
Liisa Viikari*

10.1 Introduction

Hemicellulose is somewhat loosely defined as the non-cellulose, non-pectin polysaccharide fraction of plant cell walls. In contrast to the insoluble, highly crystalline, homogeneous, and unbranched cellulose polymer, hemicellulose is more often than not highly branched, substituted, and heterogeneous, composed of a wide variety of subunits, including sugars, sugar acids, and non-carbohydrate moieties. This complex structure distinguishes the hemicelluloses from other plant cell wall polysaccharides, rendering them soluble, hygroscopic, and highly cross-linked to a variety of other plant cell wall components. Hemicelluloses are reported to be linked to lignin through cinnamate acid ester linkages, to cellulose through interchain hydrogen bonding, and to other hemicelluloses via covalent and hydrogen bonds. Though pectins have properties similar to hemicellulose, they are distinguished from the hemicelluloses by being more acidic, having longer side chains, more numerous branches, and by their lack of covalent cross-links to other cell wall components. In a general sense, hemicellulose and pectin have similar water retaining functions while differing in their cross-linking abilities. Lignin, the third major component of plant cell walls, is a large, three-dimensional polymer of various phenylpropanoid derivatives. While cellulose provides structural support and hemicelluloses cross-link and retain water, lignin acts as the overall binder, filling in gaps and providing adhesion between the various cell wall components.

The structural and compositional heterogeneity of hemicellulose spans across plant families, cell types, and even cell wall subsections. It is well established that different plant families have different ratios of the various hemicelluloses. Herbaceous plants contain mainly xylans, especially arabinoxylan, while hardwood xylans are predominantly glucuronoxylan. Softwoods, in contrast, are dominated by glucomannan, galactomannan, and galacto(gluco)mannan. In addition, other hemicelluloses are present in many plant cell walls, including xyloglucan and β -glucans (both 1 \rightarrow 3, 1 \rightarrow 4 and mixed 1 \rightarrow 3, 1 \rightarrow 4 linkages). Different hemicelluloses may occur in specific plants or as specific storage polysaccharides. Though often cited in the literature as a result of destructive compositional analysis, arabinan, galactan, and mannan occur relatively infrequently in a linear, unbranched form

(1). These terms are often used for convenience in assigning a polymeric source structure for the determined monomers. True arabinan is found in the pectin fractions of cell walls, especially beet root, fruits, leaves, and young tissues (2, 3) and as minor products released by various wood pulping and bleaching processes (4–6). Linear galactan is part of certain pectins and softwood hemicelluloses (5), and mannan occurs in ivory nut as a homogeneous linear mannose polymer. The monomer sugars (mannose, arabinose, and galactose) occur in many types of hemicellulose and pectin, so care must be used in assigning a source material for these sugars. As an example, arabinogalactan is a common hemicellulose found in larchwood. Acid hydrolysis yields mainly galactose and arabinose derived from a β -(1 \rightarrow 3)-linked galactose backbone containing galactose, arabinose, and other side groups or chain (5, 7). Some of the galactose is from the galactan backbone, some from side chains, and none of the arabinose is from arabinan. At best, it is difficult to determine the structural origin of the various monosaccharides comprising plant cell walls.

The complexities of hemicellulose extend to the enzymes involved in hydrolyzing these heterogeneous molecules. As numerous studies have pointed out, the branching and side groups or chains of many hemicelluloses affect the solubility of these compounds. Rapid debranching without backbone depolymerization results in intermolecular aggregation, causing precipitation and rendering the molecule even more difficult for depolymerization enzymes to access. Conversely, depolymerization of the backbone requires access by the depolymerases, often impeded by branches and side chains (8–10). Studies on synergy between debranching and depolymerizing enzymes have repeatedly shown that hemicellulose hydrolysis is optimal when both types are acting simultaneously (2, 11–13), Tables 10.1 and 10.2.

Given the complexities of biomass, especially that of the hemicellulose fractions, it is no wonder that plant cell walls are so recalcitrant to hydrolysis. In contrast, the amorphous characteristics of hemicelluloses make it easier to hydrolyze as compared to the crystalline cellulose. As the biomass conversion effort has developed and expanded over the past half century, many different techniques have been devised and implemented to circumvent this intrinsic resistance of cellulosic biomass to hydrolysis. Many of these have revolved around some type of chemical or thermochemical pretreatment. Examples include, among others, dilute

Table 10.1 Synergy of different xylanolytic enzymes purified from *T. reesei* in the hydrolysis of deacetylated glucuronoxylan

Enzymes	Hydrolysis products, % of d.w.		
	Methyl glucuronic acid	Reducing sugars	Xylo-oligomers (X ₁ –X ₅)
XYL	0.25	14.5	13.0
α -GLU	0.12	0.0	0.1
XYL + α -GLU	0.50	18.8	21.1
β -X	0.12	1.4	1.6
XYL + β -X	0.62	20.2	14.9
β -X + α -GLU	1.12	5.8	5.3
XYL + β -X + α -GLU	4.88	33.5	28.8

Abbreviations: α -GLU, α -glucuronidase 740 nkat/g; β -X, β -xylosidase 1000 nkat/g; XYL, xylanase pl 5.5 10 000 nkat/g, hydrolysis 40°C for 24 hours (64).

Table 10.2 Hydrolysis of steamed birchwood hemicellulose fraction by different xylanolytic enzymes from *T. reesei*

Enzymes	Hydrolysis products, % of d.w.		
	Acetic acid	Xylose	Xylobiose
XYL	4.5	6	6.0
β -X	0.5	6	0
AE	1.2	0	0
XYL + β -X	5.9	25	1.0
XYL + AE	8.5	8	16.0
β -XYL + AE	3.2	16.0	0
XYL + β -X + AE	9.4	42.0	1.0
XYL + β -X + AE + α -GLU	12.0	50.0	3.0

Abbreviations: XYL, xylanase 12 000 nkat/g; β -X, β -xylosidase 500 nkat/g; AE, acetyl esterase 500 nkat/g; α -GLU, α -glucuronidase 20 nkat/g; hydrolysis 45°C for 24 hours (77).

sulfuric acid under elevated temperature and pressure, high pressure–high temperature liquid hot water, alkaline peroxide, calcium oxide (lime), aqueous ammonium, organosolv pulping, clean fractionation, and concentrated acid. Ammonia fiber explosion/expansion (AFEX) and steam explosion (with or without acid catalyst) are variations that utilize the explosive decompression of a gaseous phase impregnated into the cell wall under pressure when the material is flashed to atmospheric pressure. The end result of these varying pretreatments is a biomass that has been rendered more digestible by enzymes; however, there are issues with this approach. Degradation products, enzyme and microbe inhibitor formation, large capital or operating expenses, and non-optimal digestibilities are some of the problems faced by pretreatment.

Historically, pretreatments have been designed and optimized to make more digestible cellulose, without considering hemicellulose digestion or hydrolysis as a factor. However, some hemicellulose modification takes place during all pretreatments; either solubilization, structural modification, or some degree of both. Both acidic and alkaline conditions usually lead to partial solubilization of hemicellulose, the degree of which is dependent upon the severity. Depending upon the pH, debranching of both soluble and residual solid fractions is a common occurrence, with alkaline conditions removing ester-linked side chains, such as acetyl and feruloyl side groups, and acidic conditions hydrolyzing arabinofuranosyl, 4-*O*-methylglucuronosyl, and other glycosidic side group linkages. More recently, in an attempt to reduce the costs and losses associated with these various pretreatments, the processes have been decreased in severity, which does help the economics of the pretreatment, but often results in less digestible biomass, mostly associated with the decrease in hemicellulose removal or hydrolysis. Now, in an attempt to maintain the advantages of reduced severity pretreatment while continuing to increase conversion rates and levels, hemicellulases are becoming much more prominent in biomass conversion technology. Today, however, there are only a few reports on the role and mechanism of enzymatic hydrolysis of residual hemicellulose enhancing the total hydrolysis of lignocellulosic residues.

Taking into account the innate complexities of hemicellulose, the variability of hemicellulose types, and the wide range of resultant biomass “products” produced by the gamut of pretreatment chemistries, enzymatic hydrolysis of hemicellulose becomes a challenge (rather than a problem). In order to find the optimal set of enzymes to digest a specific biomass, the source biomass, pretreatment chemistry, and desired end products must all be considered. The resultant products of the numerous pretreatment chemistries are widely varied and information on these can be found in other sections of this book. In nature, these complex macromolecules are completely hydrolyzed, though the process may take years, multiple organisms, and a very complex suite of enzymes. In essence, pretreatment is a much more rapid and direct substitute for some or most of these natural factors, enabling a much faster conversion, though subject to the losses described earlier. Depending on the pretreatment selected, the optimum mix of enzymes for a given conversion process will presumably be a subset of the complete natural digestion suite. Some activities may no longer be needed at all (potentially esterases in an alkaline process) or needed in lower amounts. The key is determining which required activities have changed as a result of pretreatment.

10.2 Hemicellulase types, activities, and specificities

The term *hemicellulase* is often used to refer to a mix of enzyme activities that act upon the non-cellulose, non-pectin polysaccharides in biomass. The diversity and complexity of plant cell wall structural hemicelluloses preclude a more precise definition. In general terms, the enzymes involved in hemicellulose degradation can be divided into two major categories: depolymerizing and debranching. Tables 10.3–10.6 have been compiled from several online enzyme databases, the Carbohydrate Active eZyme database (CAZy, <http://www.cazy.org>), the Expert Protein Analysis System (ExPASy, <http://www.expasy.org>), and BRENDA, the Comprehensive Enzyme Information System (<http://www.brenda.uni-koeln.de>) (14–16). These tables further categorize the enzymes involved in cellulose and hemicellulose degradation and demonstrate the complexity of the problem. It is important to understand that the classification of biomass-degrading enzymes spans several nomenclatures, with some, but nowhere near complete, cross-referencing. One common system is the IUMB system, with the enzymes being classified by four numbers, designating the enzyme type and activity, i.e., 3.2.1.4. The second system, which has gained much popularity over the last decade or so, is the so-called CAZy database, which, in contrast to the IUMB system, classifies glycosyl hydrolases (and other carbohydrate-active enzymes) into families based on structural and evolutionary similarities, GH7 for example. As there is no comprehensive and simple cross-reference available between these two systems, enzyme designations here are given in either the original connotation from the referenced material, or in the current most common usage.

Depolymerizing enzymes act on the backbone sugar chain and are usually classified as either endo-acting, which cut the chain in the midst of a long polymer, or exo-acting, which work from the end of the chain. Several, however, have been reported to be associated with both types of activity. In addition, there are a series of enzymes that act on the oligomers

Table 10.3 Cell wall polysaccharide depolymerizing β -glucanases

IUPAC	Name	Families	Dominant substrates	Dominant products	Dominant linkages
3.2.1.4	Cellulase	5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 61, 74	Cellulose, β -glucans, mixed linkage β -glucans	Glucan oligomers	(1 \rightarrow 4)- β -D-glucoside
3.2.1.6	Endo-1,3(4)- β -glucanase	16	Laminarin, lichenin, mixed β -glucans	Glucan oligomers	(1 \rightarrow 3,4)- β -D-glucoside
3.2.1.21	β -glucosidase	1, 3, 9	Cellobiose, other oligoglycans	Glucose	(1 \rightarrow 4)- β -D-glucoside
3.2.1.39	Glucan endo-1,3- β -glucosidase	15, 17, 55, 64, 81	Laminarin	Glucan oligomers	(1 \rightarrow 3)- β -D-glucoside
3.2.1.58	Glucan 1,3- β -glucosidase	3, 5, 17, 55	Laminarin	Glucose	Non-reducing end (1v3)- β -D-glucoside
3.2.1.73	Licheninase	5, 8, 11, 12, 16, 17	Lichenin, mixed β -(1 \rightarrow 3,4) glucans	Mixed β -(1 \rightarrow 3,4)-glucans	(1 \rightarrow 4)- β -D-glucoside
3.2.1.74	Glucan 1,4- β -glucosidase	3	Cello-oligomers	Glucose	exo (1 \rightarrow 4)- β -D-glucoside
3.2.1.91	Cellulose 1,4- β -cellobiosidase	5, 6, 7, 9	Cellulose	Cellobiose	Non-reducing end β -(1v4) glucoside
3.2.1.120	Oligoxyloglucan β -glycosidase		Xyloglucan oligomers	Isoprimeverose (xyl-p- α -(1 \rightarrow 6)- β -D-gluc-p)	(1 \rightarrow 4)- β -D-glucoside
3.2.1.150	Oligoxyloglucan reducing-end-specific cellobiohydrolase	74	Xyloglucan oligomers	Cellobiose-non-reducing glc may be substituted xyl-p- α -(1 \rightarrow 6)	Reducing end (1 \rightarrow 4)- β -D-glucoside
3.2.1.151	Xyloglucan-specific endo- β -1,4-glucanase	5, 12, 16, 26, 44, 74	Xyloglucan	Xyloglucan oligomers	(1 \rightarrow 4)- β -D-glucoside
3.2.1.155	Xyloglucan-specific exo- β -1,4-glucanase		Xyloglucan	Xyloglucan oligomers	(1 \rightarrow 4)- β -D-glucoside, mixed endo/exo modes

Table 10.4 Cell wall depolymerizing β -xylanases

IUPAC	Name	Families	Dominant substrates	Dominant products	Dominant linkages
3.2.1.8	Endo-1,4- β -xylanase	5, 8, 10, 11, 16, 43, 62	Xylan	Xylan oligomers (may still have side chains)	(1 \rightarrow 4)- β -D-xyloside
3.2.1.32	Xylan endo-1,3- β -xylosidase	10, 26	β -(1 \rightarrow 3)-linked Xylans	β -(1 \rightarrow 3) Xylan oligomers	(1 \rightarrow 3)- β -D-xylosyl
3.2.1.37	Xylan 1,4- β -xylosidase	3, 30, 39, 43, 51, 52, 54	Xylan	Xylose	Reducing end (1 \rightarrow 4)- β -D xylosyl
3.2.1.72	Xylan 1,3- β -xylosidase		β -(1 \rightarrow 3)-linked Xylans	Xylose	Non-reducing end (1 \rightarrow 3)- β -D xylosyl
3.2.1.136	Glucuronoarabinoxylan endo-1,4- β -xylanase	5	Glucurono-feraxan (ferulated arabinoxylan)	Xylan	(1 \rightarrow 4)- β -D-xylosyl adjacent to glucuronosyl substituted xylose
3.2.1.156	Oligosaccharide reducing-end xylanase	8	Xylo-oligomers	Xylose	Reducing end (1 \rightarrow 4)- β -D xylosyl

Table 10.5 Other cell wall depolymerizing glycosyl hydrolases

IUPAC	Name	Families	Dominant substrates	Dominant products	Dominant linkages
3.2.1.25	β -mannosidase	1, 2, 5	Mannan, manno-oligomers	Mannose	Non-reducing end (1 \rightarrow 4)- β -D-mannosyl
3.2.1.78	Mannan endo-1,4- β -mannosidase	5, 26, 4	Mannan, glucomannan galactomannan	Manno-oligomers	(1 \rightarrow 4)- β -D-mannosyl
3.2.1.89	Arabinogalactan endo-1,4- β -galactosidase	53	Arabinogalactan, pectin	Galacto-oligomers	(1 \rightarrow 4)- β -D-galactosyl
3.2.1.99	Arabinan endo-1,5- α -L-arabinosidase	43	Linear arabinan	Arabino-oligomers	(1 \rightarrow 5)- α -L-arabinosyl
3.2.1.100	Mannan 1,4-mannobiosidase		Mannan	Mannobiose	Non-reducing end (1 \rightarrow 4)- β -D-mannosyl
3.2.1.145	Galactan 1,3- β -galactosidase	43	Arabinogalactan	Galactose	Non-reducing end (1 \rightarrow 3)- β -D-galactosyl

Table 10.6 Cell wall polysaccharide debranching enzymes

IUPAC	Name	Families	Dominant substrates	Dominant products	Dominant linkages
3.1.1.72	Acetylxylan esterase		O-Acetylated xylan	Acetic acid	Xylose- <i>O</i> -acetyl
3.1.1.73	Ferulic acid esterase		Ferulated xylan	Ferulic acid	Arabinose- <i>O</i> -feruloyl
3.1.1.6	Acetyl esterase		O-acetylated xylan/xylo-oligomers	Acetic acid	Xylose- <i>O</i> -acetyl
3.2.1.131	Xylan α -1,2-glucuronosidase	67	(4- <i>O</i> -methyl)-glucuronoxylan	Glucuronic acid, 4- <i>O</i> -methyl glucuronic acid	(1 \rightarrow 2)- α -D-glucuronosyl
3.2.1.139	α -glucuronidase	4, 67	(4- <i>O</i> -methyl)-glucuronoxylan	Glucuronic acid, 4- <i>O</i> -methyl glucuronic acid	(1 \rightarrow 2)- α -D-glucuronosyl
3.2.1.55	α - <i>N</i> -arabinofuranosidase	3, 10, 43, 51, 54, 62	Arabinan, arabinogalactan, arabinoxylan	Arabinose	(1 \rightarrow 3,5)- α -L-arabinosyl

generated by the combination of endo- and exo-activity. These include β -glucosidase (3.2.1.21), β -xylosidase (3.2.1.37), and β -mannosidases (3.2.1.25) among others. Due to the low degree of polymerization of their substrates, distinction between endo- and exo-hydrolysis modes of action is difficult.

Debranching enzymes, often referred to as accessory enzymes, can be further subdivided into those acting on glycosidic linkages and those acting on ester-linkages. The dominant enzymes of the former include α -L-arabinofuranosidases (3.2.1.55) and α -glucuronidases (3.2.1.139), both of which remove glycosidic side chains from xylan (2, 13). The dominant esterases include acetyl xylan esterase (3.1.1.72) and feruloyl esterase (3.1.1.73), both of which also act on xylan (11). Enzymes acting on the major hemicelluloses are described in Figure 10.1. There are also reports of other esterases acting on other acetylated polysaccharides, including glucomannan, galactomannan, and even cellulose (12).

Some hemicellulases and accessory enzymes exhibit cross reactivity across the different hemicelluloses, while others may be very specific for a particular oligomers sequence or conformation. β -xylosidase, for example, may preferentially hydrolyze xylobiose, but may also act on xylotriose and higher xylo-oligomers, gentibiose, and cellobiose. Feruloyl esterase, though normally most active on ferulic acid ester-linked to arabinose, may also be active on coumaroyl esters. β -glucosidases have a wide range of specificities, with certain enzymes acting across a broad range of β -(1 \rightarrow 4), β -(1 \rightarrow 3), and mixed β -(1 \rightarrow 3,4) linkages and other versions of the same enzymes requiring a specific linkage sequence or side chain pattern. Due to the high degree of heterogeneity in hemicellulose structure, the number of permutations involved in the conformation of sugars, non-sugars, and their linkages

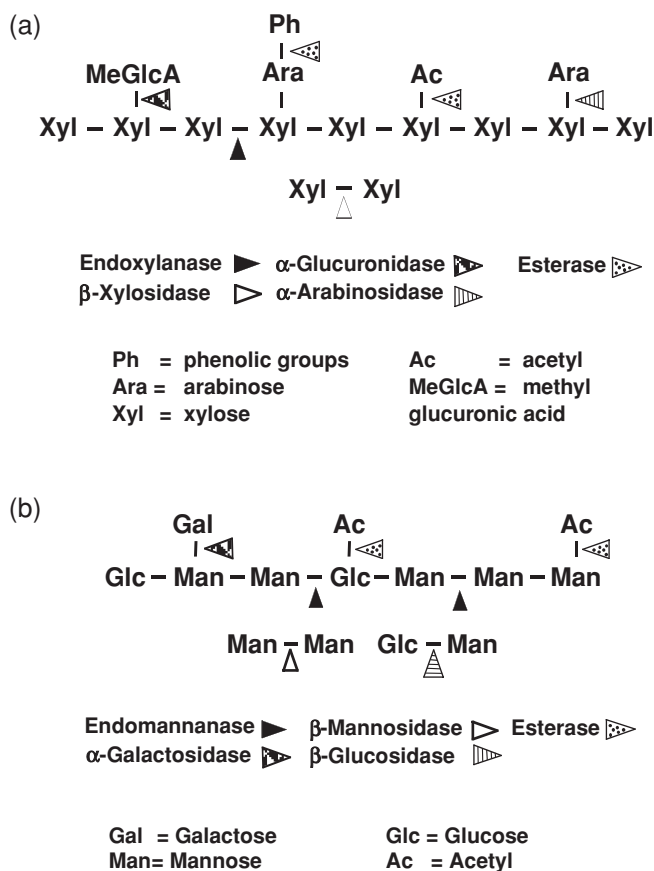


Figure 10.1 Enzymes participating in the hydrolysis of xylans (a) and glucomannans (b): Ac, acetyl; Ph, phenolic groups; Ara, arabinose; MeGlcA, methyl glucuronic acid; Xyl, xylose; Gal, galactose; Glc, glucose; Man, mannose.

precludes a comprehensive, or even in-depth, discussion of the specific enzymes and their activities. Instead, we will outline the basic activities of a variety of hemicellulase enzymes and indicate their potential use in the depolymerization of the key bioenergy feedstock hemicelluloses.

10.3 Depolymerases

10.3.1 Xylanases

Xylans from different plant sources are categorized by the substituents of their side chains (i.e., arabinoxylan, glucuronxylan, etc.); however, all xylans are built upon a poly- β -(1 \rightarrow 4)-xylopyranose backbone. Xylanases depolymerize the β -(1 \rightarrow 4)-xylopyranose backbone of

a variety of xylans, with the endoxylanases classified in 3.2.1.8 [β -(1 \rightarrow 4) hydrolysis] or 3.2.1.32 [β -(1 \rightarrow 4) or β -(1 \rightarrow 3) hydrolysis]. In both groups, the anomeric configuration of the reducing-end xylan is retained. Most xylanases belong to the two structurally different glycosyl hydrolase groups (families 10 and 11) and differ from each other with respect to their catalytic properties (17). Xylanases with high Mr/low pI (family 10) seem to exhibit greater catalytic versatility than the low Mr/high pI-xylanases (family 11) and thus they are, for example, able to more efficiently hydrolyze highly substituted xylans. Few xylanases have been classified in family 5, which is a very versatile family containing various types of glycanases. Some xylanases have been reported to contain either a xylan-binding domain (18) or a cellulose-binding domain (19). Some of the binding domains have been found to increase the degree of hydrolysis of fiber-bound xylan, whereas others have no effect thereon.

Most of the xylanases characterized are able to hydrolyze different types of xylans, showing only differences in the spectrum of end products. However, no systematic studies on the substrate specificity of xylanases belonging to different families on fiber-bound substrates have been carried out. The majority of the endoxylanases produce mainly xylobiose, xylotriose, and substituted oligomers of two to four xylosyl residues as products upon extended incubation. Xylose, xylopentaose, and higher oligomers may also be produced, with specific product patterns being dependent upon the individual enzyme and the substrate. Most of the endoxylanases hydrolyze unsubstituted xylose polymers, with the tolerance of side chains again being dependent on the specific xylanase.

The exoxylanases are found in 3.2.1.37, 3.2.1.72, and 3.2.1.156, though enzymes in 3.2.1.72 are active on β -(1 \rightarrow 3)-linked xylosides and have limited activity on the majority of xylans. The other two categories, though similar in activity, are distinguished by their difference in preferred substrate degree of polymerization, with the β -xylosidase (1,4- β -D-xyloside xylohydrolase, EC 3.2.1.37) acting more effectively on polymeric xylan, while enzymes in 3.2.1.156 have a greater affinity for xylo-oligomers. Both groups act at the reducing end of the chain, yielding xylose as their product. Exoxylanases are less abundant and consequently less well known and understood than the endoxylanases, as only a few examples of the former have been characterized. Taking into account the high degree of substitution of native xylans, the exo-mode of action may be less dominant among hemicellulases than cellulases. Exoglycanases are generally larger proteins than endoglycanases, with molecular weights above 100 kDa and they are often built up by two or more subunits. The β -xylosidases, which have a few well-studied examples and though active primarily on xylobiose, may have a processive activity on xylo-oligomers.

10.3.2 Mannanases

Endomannanases (EC 3.2.1.78) catalyze the random hydrolysis of β -D-1,4 mannopyranosyl linkages within the main chain of mannans and various polysaccharides consisting mainly of mannose, such as glucomannans and galactoglucomannans. Far fewer mannanases are characterized as compared with the numerous xylanases. The mannanase of *Trichoderma reesei* has a similar multi-domain structure to several cellulolytic enzymes; i.e., the protein contains a catalytic core domain which is connected by a linker to a cellulose-binding domain

(20, 21). The CBD increases the action of *T. reesei* mannanase on fiber-bound glucomannan, even though the catalytic domain can efficiently degrade crystalline mannan (22).

The main hydrolysis products from galactoglucomannans and glucomannans are manno-biose, mannotriose, and various mixed oligosaccharides. The hydrolysis yield is dependent on the degree of substitution as well as on the distribution of the substituents (23). The hydrolysis of glucomannans is also affected by the glucose to mannose ratio. Some mannanases are able to hydrolyze not only the β -1,4-bond between two mannose units, but also the bond between the adjacent glucose and mannose units (24, 25). Interestingly, some mannanases are able to degrade mannan crystals quite efficiently (26, 27).

Enzymes needed for further hydrolysis of the released manno-oligomers produced by endoenzymes are β -mannosidase (1,4- β -D-mannoside mannohydrolase, EC 3.1.1.25) and β -glucosidase (EC 3.2.1.21). β -mannosidase and β -glucosidase catalyze the hydrolysis of manno-oligosaccharides by removing successive mannose or glucose residues from the non-reducing termini. The β -xylosidase of *T. reesei* and β -mannosidases of *A. niger* are also able to attack polymeric xylan and mannan, respectively, liberating xylose and mannose by successive exo-action (28, 29).

10.3.3 β -glucanases

Endoglucanases are most often referred to in the context of cellulases. Though these enzymes have a high affinity toward cellulose, they often exhibit cross reactivity with other β -glucans, including xyloglucan and mixed β -(1 \rightarrow 3, 1 \rightarrow 4)-glucans. As with the endoxylanases, these enzymes cleave β -(1 \rightarrow 4) or β -(1 \rightarrow 3) linkages in the interior of the glucose chain, generating two new chain ends, one reducing and one non-reducing. Different enzymes are required to cleave the two forms of β -glucan (28, 30–35). Enzymatic degradation of β -glucan is accomplished through glycosyl hydrolase family 12 enzymes (EC 3.2.1.4). Although these endo-acting enzymes are active on β -(1 \rightarrow 4) glycosidic linkages, they are differentiated from other β -(1 \rightarrow 4)-acting enzymes by being able to hydrolyze the β -(1 \rightarrow 4) linkages in mixed β -(1 \rightarrow 3, 1 \rightarrow 4)-linked polysaccharides. Glucan endo-1,3- β -D-glucosidase [β -(1 \rightarrow 3) glucanase] (EC 3.2.1.39) is an endo-acting glycosyl hydrolase that acts on β -(1 \rightarrow 3) glucan, but has very limited activity on the mixed linkage β -glucan. Endo-1,3(4)- β -glucanase [β -(1 \rightarrow 3, 1 \rightarrow 4) glucanase] (EC 3.2.1.6), is also an endo-acting glycosyl hydrolase. There is an exo-acting glycosyl hydrolase that is active on β -(1 \rightarrow 3) glucan. Glucan 1,3- β -glucosidase (EC 3.2.1.58) acts by processively releasing glucose from β -(1 \rightarrow 3) glucan from the non-reducing end.

Much of the initial work on cellulases and endoglucanases was concentrated on the enzymology of the fungus *Trichoderma reesei*. This fungus produces several cellulases which act synergistically in the degradation of cellulose. In *T. reesei*, the Cel7B is a major endoglucanase, forming about 6–10% of the total *T. reesei* cellulase (36, 37). It has wide activity against solid and soluble substrates, such as CMC, as well as against xylan and glucomannans (38). Also, the endoglucanase Cel5A has activity against solid (celluloses) and soluble (CMC, mannan) substrates (39–41), but not on xylans. This enzyme comprises up to 10% of the total cellulases in *T. reesei* (36, 37). The minor endoglucanases (Cel12A, Cel45A) are reported to hydrolyze solid and soluble (including glucomannan) substrates with diverse specific activities.

10.3.4 Xyloglucanases

Xyloglucan is one of the major hemicellulose polymers in growing primary cell walls of various plant species. Due to analytical problems, it has been difficult to differentiate it from cellulose and xylan. Xyloglucan is closely associated with cellulose microfibrils via hydrogen bonding, thus providing the load-bearing network of the cell wall, which protects the cell wall from collapsing due to the osmotic stress. Most research efforts have been focused on determination of plant enzymes responsible for control and modification of the expanding cell wall (42–60). Although not as heavily branched as xylans, the xylose and other substituents on xyloglucan can make enzymatic digestion more complicated than that of cellulose and β -glucan.

Xyloglucan has cellulose-like backbone of β -1,4-D-glucopyranose residues to which α -D-xylose residues are attached at C-6 position. Generally, 60–75% of the glucose residues are branched with xylose, except in grasses, which have a lower degree of substitution. Xylose can form side chains with D-galactopyranose and L-fucopyranose saccharides, and rarely with L-arabinofuranose. Plant cell wall degrading enzymes, such as endoglucanases, xyloglucan endotransglycosylases, and exoglycosidases, like α -fucosidases or β -galactosidases have been reported to digest the xyloglucan. Also, some cellulases of *Trichoderma reesei* are able to hydrolyze the xyloglucan backbone (61). A new class of polysaccharide-degrading enzymes comprise the specific xyloglucanases, or xyloglucan-specific endoglucanases, which can attack the backbone also at substituted glucose residues (62).

Some xyloglucanases require a specific xylose substitution pattern while others are more general (43). This determination seems to be dependent on the binding sub-sites in specific endoglucanases (61). A xyloglucanase from *Aspergillus niger* has been shown to be active against several β -glucans, but having the highest activity against tamarind xyloglucan (63). This, combined with its lack of synergy with cellulases, indicates specificity different from traditional endoglucanases. A plant-specific enzyme believed to be responsible for modification of xyloglucan in the cell wall through endo-hydrolysis and glycosyl transferase activities has also been characterized (59, 60).

10.4 Debranching enzymes (accessory enzymes)

Glycosidic side groups connected to xylan and glucomannan main chains are primarily removed by α -glucuronidase, α -arabinofuranosidase, and α -D-galactosidase. Acetyl and hydroxycinnamic acid substituents bound to hemicellulose are removed by acetyl xylan esterases and ferulic/coumaric acid esterases (Figure 10.1). These are clearly different types of side-group cleaving enzymes. Some of them are able to hydrolyze only substituted short-chain oligomers, which first must be produced by the backbone depolymerizing endoenzymes (xylanases and mannanases). Others are capable of debranching intact polymeric substrates. Most accessory enzymes of the latter type, however, prefer oligomeric substrates. The synergism between different hemicellulolytic enzymes is also observed by the accelerated action of endoglucanases in the presence of accessory enzymes. Side groups that are still attached to oligosaccharides after the hydrolysis of xylans and mannans by xylanase or mannanase, respectively, restrict the action of β -xylosidase and β -mannosidase.

10.4.1 α -glucuronidase

α -glucuronidase (3.2.1.139) catalyzes the release of glucuronic acid or 4-*O*-methylglucuronic acid from xylan. Reports of specific substrate specificities vary from activity on long-chain xylans to a requirement for attachment to a terminal non-reducing end xylose (64, 65). At least one membrane-bound bacterial enzyme exhibited activity specifically on soluble xylan-derived oligomers (66). However, synergy between α -glucuronidase and endoxylanase on wheat xylan has been shown to be significant with simultaneous activity yielding the highest release of 4-*O*-methylglucuronic acid (67).

10.4.2 α -arabinofuranosidase

α -arabinofuranosidase cleaves arabinose side chains from the xylan main chain. Many enzymes in this class have demonstrated activity on pectin, arabinan, and arabinoxylan with the preferred chain length and side-group assignments also varied (13, 68, 69). A comprehensive review can be found here (69). Synergy with xylanases has been reported, as has synergy with ferulic acid and acetyl xylan esterase (2, 13, 65, 67, 70). As ferulic acid is linked to the arabinose, which in turn is linked to the xylan chain, these synergies are not unexpected. The substrate specificity of this enzyme class is still somewhat muddled (65).

10.4.3 α -D-galactosidase

α -D-galactosidase activity is required for hydrolysis of softwood mannans, specifically galactomannans and galactoglucomannans (71). The enzyme acts on the α -galactosyl side groups attached to the O-6 position of the backbone mannose units (65). Little work has been collected regarding this enzyme, though its importance in softwood pulping has been considered.

10.4.4 Acetyl xylan esterase

Acetyl groups occur on several hemicelluloses although the primary examples are xylan and galactoglucomannans. Cereal and hardwood xylans have much higher levels of acetylation than softwood xylans. Softwood acetylation occurs principally in the galactoglucomannans. The most likely reason for acetylation is to keep the hemicelluloses soluble and hydrated. Deacetylation of xylan and glucogalactomannan results in significantly decreased solubility of the polymer. Acetyl groups also cause problems for microorganisms when they are released from the main chain, resulting in decreased pH. The release of acetate is inhibitory to many microbes and is a considerable problem in the conversion of biomass to fermentation products (72–75).

Acetyl groups are released from hemicelluloses both from high degree of polymerization (i.e., native) substrates and from acetylated oligomers resulting from depolymerization. Acetyl xylan esterases (AXEs) may exhibit a preference for one or the other form, or may act on both types (76). Synergy studies between AXE, xylanases, and other hemicellulase

enzymes have repeatedly demonstrated that the most effect digestion occurs with the appropriate ratios of all enzymes acting simultaneously (70, 76–78). Debranching in the absence of depolymerization results in insoluble long-chain hemicelluloses that can be more difficult for the depolymerases to access (78). Depolymerization without debranching may be limited by depolymerase access to the polymer main chain. Most AXEs are low molecular weight and may or may not have a carbohydrate-binding module.

10.4.5 Ferulic acid esterase

Ferulate side chains are found in cereal and hardwood xylans, as well as in many types of pectin. In xylan, they are ester-linked to the C-2 position of the arabinose side chain of the xylan backbone, where they function as crosslinkers through ether linkages to either ferulic acid on another xylan chain or to lignin components (79). This provides some three-dimensional stability to the polymer network (80). Ferulic acid esterases (FAEs) are active in cleaving ferulic acid from these polymers, though the specificity is again not clear. Some have been reported to act on coumaric as well as ferulic acid (81). Some are preferentially active on polymers while others are more active on substituted xylo-oligomers. There are also reports of FAEs having activity on both ferulated xylan and ferulated pectin (82). Atomic force microscopy studies have suggested that hydrolysis of ferulic acid bridges results in shorter, less-branched xylan chains (83). Other studies have reinforced the synergy between xylanase and FAE, including enhanced synergy in an FAE/xylanase fusion protein (11, 67, 84–86).

10.5 Hemicellulase activities for biomass feedstocks

Until recently, pretreatments have usually been designed for extensive removal of hemicellulose in order to improve the enzymatic hydrolysis of cellulose. Thus, the impact of the enzymatic hydrolysis of fiber-bound hemicellulose on cellulose hydrolysis has not been considered important, as evaluated by the number of publications. The amount of hemicellulose in the solid substrate resulting from pretreatment varies from about 1% to up to 25%, depending upon the pretreatment. Acidic and high temperature pretreatments tend to hydrolyze and remove more hemicellulose, while alkaline and low temperature processes generally leave higher residual hemicellulose in the solids and greater amounts of oligomers in the hydrolyzates. Ammonium fiber expansion/explosion (AFEX) is essentially a dry process and results in virtually no change in the solids composition, as there is no liquid phase to partition components.

The removal of xylans during the pretreatment has been shown to correlate with the hydrolyzability of the raw material (87). There was a clear correlation with the residual hemicellulose content of pretreated spruce and the degree of hydrolysis. It is however, difficult to conclude whether this correlation is affected also by other factors, such as further chemical modifications caused by the severity of the pretreatment. In another study, various raw materials with different levels of residual hemicellulose contents did not seem to follow this hypothesis. Neither residual xylan nor glucomannan seemed to correlate with the degree of cellulose hydrolysis when additional β -glucosidase was supplemented. On the other hand,

xylanase activity in the preparation was shown to increase the hydrolysis conversion and rate (88).

Hemicellulolytic activity in commercial cellulase preparations has been expected to provide the necessary hydrolysis of the residual hemicellulose in the solid matrix. The activities of accessory enzymes in the commercial preparations vary, and in most reports, are not even measured. Their role in the solubilization of matrix bound hemicelluloses can only be speculated. Even less is known about the effect of these unquantified activities on the soluble oligomer fractions resulting from pretreatment. Although extensive studies on the activities required for enzymatic hydrolysis of these compounds have not been carried out, interest and research into this area are rapidly expanding. From the known structures and known enzyme activities, one can begin to evaluate enzyme mixes for their efficacy on these complex substrates.

10.5.1 Xylan

The complexity of plant cell wall xylans implies a correlative complexity in the enzymes needed for digestion. Further complexity is imparted by pretreatment conditions and severity. Xylan composition varies by type, with softwood xylan being the most distinct from hardwood and herbaceous xylans. Different debranching enzymes are required depending on the specific type of xylan being hydrolyzed. These include arabinofuranosidases, ferulic and coumaric acid esterases, acetyl and acetyl xylan esterases, glucuronidases, and xylosidases. Simultaneous removal of these side chains and cleavage of the polymeric backbone synergistically enhances the rate of degradation by endoxylanase enzymes (67, 76, 78, 89–93). Thus, a xylan that has been subjected to acetyl xylan esterase is less susceptible to enzymatic degradation than a xylan subjected to a mixture of depolymerizing and debranching enzymes (69, 77).

The xylan of softwoods is, like all xylans, composed of a β -(1 \rightarrow 4)-D-xylopyranose backbone (65). In softwoods, the side chains are dominated by α -(1 \rightarrow 2)-linked 4-O-methylglucuronic acid. The reported uronic acid:xylose content varies with species and extraction method, ranging from 2:10 up to 7:10 glucuronic acid:xylose. α -L-arabinofuranose units are linked (1 \rightarrow 3) to the backbone as well, although only one arabinose per every eight or nine xyloses is typical. Unlike hardwood and herbaceous xylans, softwood xylans contain no acetyl groups (65). For softwood xylans, the dominant debranching activities are likely to include α -glucuronidase and α -L-arabinofuranosidase, with the lack of acetyl groups eliminating the need for acetyl esterases. Acetyl esterases, however, may still be required for depolymerization of the galactoglucomannans. The limited arabinose substitution also limits the presence of ferulic acid, though there may be a low requirement for FAE.

Hardwood xylan is composed of β -D-xylopyranosyl units which may contain 4-O-methyl- α -D-glucuronic acid and acetyl side groups. 4-O-Methylglucuronic acid is linked to the xylan backbone by O-(1,2) glycosidic bonds and the acetic acid is esterified at the C2 and/or C3 hydroxyl group. Typically, the polymer is debranched, either prior to or simultaneously with depolymerization of the backbone. As the substituents are removed, xylan can become less soluble and form aggregates that sterically hinder and retard further degradation (78). In hardwood and cereal xylans, debranching activity requirements are

predominantly acetyl esterases and arabinofuranosidases, as well as ferulic acid esterases. α -glucuronidases are needed in a more limited capacity compared to softwood xylans.

Since the backbone for all xylans is essentially identical, the basic depolymerase activities are essentially the same regardless of feedstock; endo- and exoxylanases with β -xylosidase activity for oligomer hydrolysis. The specific enzyme requirements may be quite different, however, as the structural and side chain requirements of the depolymerases are often dependent upon the residual side groups.

The endoxylanases randomly cleave the 1,4- β -D-xylosidic linkages of the main chain and are often quite particular about the type of linkage, type of sugar, and presence or absence of nearby substituents (77). Although there are such specific examples of endoxylanases requiring side chains for maximal activity (94), the majority of the endo-acting hemicellulose hydrolases tend to be more active on debranched or partially debranched hemicellulose, especially in the case of xylanases. A balance between removing the side groups from the polysaccharide backbone, decreasing the average chain-length, and hydrolyzing the oligomers into free monomers must be met. The concerted action of the various hemicellulase enzyme classes probably accounts for the high synergy observed when the enzymes are used concert with each other (95).

10.5.2 *Galactoglucomannan and glucomannan*

The dominant hemicelluloses in softwoods are glucomannan and galactoglucomannan (GGM) (65). The GGM structure varies by species and cell wall location, but is generally built upon a β -(1 \rightarrow 4)-D-mannopyranose and β -(1 \rightarrow 4)-D-glucopyranose backbone. The ratio and sequence of the two backbone sugars vary, but typically several mannose units are linked together and periodically interrupted by a glucose unit. The usual man:glc ratio is 3:1. Side chains are dominated by α -D-galactopyranose units, typically linked to the O-6 position of either backbone sugar, though O-2 and O-3 linkages on the mannose units have been reported for some species. The galactose side units are typically single, but may be double, with the second terminal galactose linked β -(1 \rightarrow 2) to the first unit. The other common substituent is an ester-linked acetyl group. Typically, these are linked at the O-2 or O-3 positions of the backbone sugars and reported content and substitution patterns vary widely. As with acetylated xylans, one of the primary effects of acetyl content is the effect on the solubility of these polymers in water. High acetyl content GMMs tend to be water soluble, as do GMMs with high galactose content. The alkali-soluble GMMs appear to have fewer side chains, though the extraction conditions remove any ester-linked acetyl groups by saponification. The ratio of galactose to glucose is often much lower in these as well and they are frequently referred to as glucomannans (65).

Despite a lack of systematic studies on enzymatic degradation of softwood, much can be deduced from the structural information regarding the enzyme activities necessary for hydrolysis. Acetyl esterase and α -galactosidase are the dominant debranching enzymes, with the requirement for either being dictated by the specific type of softwood glucomannan or galactoglucomannans being hydrolyzed (12, 76). Endomannanases and β -glucanases are the dominant main-chain depolymerases and β -mannosidases and β -glucosidases are likely needed for reduction of the oligomers to monomer sugars.

10.5.3 Arabinogalactan, xyloglucan, and β -glucan

Other minor hemicelluloses found in biomass feedstocks include arabinogalactan, xyloglucan, and β -glucans. Arabinogalactan, found mainly in softwoods, is comprised of a linear β -(1 \rightarrow 3)-D-galactopyranose backbone highly substituted at the C-6 position, though β -(1 \rightarrow 4) galactans have also been identified (5, 96). These side chains include β -D-galactose, α -L-arabinose, and β -D-glucuronic acid in either monomer or dimer configurations. From the configuration, debranching with β -galactosidases, α -L-arabinofuranosidase, and β -glucuronidases is likely to be supplemental to depolymerization by β -galactanases.

Xyloglucan is a linear β -(1 \rightarrow 4)-D-glucopyranose polymer with α -(1 \rightarrow 6)-D-xylopyranose side chains, typically present as a monomer, but also extended by (1 \rightarrow 2) linkages to β -galactose which may be further extended by (1 \rightarrow 2) α -fucose. Specific xyloglucanases which depolymerize the β -(1 \rightarrow 4)-D-glucopyranosyl backbone have been identified and most fall into the GH74 family (62, 63, 97). The distinction between these endoxyloglucanases and endocellulases may lie in the requirement for specific side groups in the case of the xyloglucanases and synergy has been demonstrated between these enzymes (61, 62, 98, 99). Little work has been done on the debranching enzyme requirements in xyloglucan, though α -xylosidases are obviously important candidates.

The β -glucans of softwoods are typically β -(1 \rightarrow 3)-D-glucopyranose chains with a few β -(1 \rightarrow 4) linkages and the rare glucuronic or galacturonic side chain. Other potential, though minor, β -glucans from biomass feedstocks include polymers such as endosperm β -glucan, a mixed (1 \rightarrow 3,4)- β -glucan, where the (1 \rightarrow 4) linkages are periodically replaced every three or four residues by a (1 \rightarrow 3) linkage. As with cellulose, endo- and exoglucanases are the major contributors, however, the specificity of the enzymes may be to a specific linkage or series of linkages. The so-called lichenanases (3.2.1.73) hydrolyze the β -(1 \rightarrow 4) bond but require an adjacent 3-O-substituted glucose and will not hydrolyze homogeneous β -(1 \rightarrow 4) glucans (100). Laminarinases (3.2.1.39) hydrolyze β -(1 \rightarrow 3) bonds. Homogeneous stretches of β -(1 \rightarrow 4) links are hydrolyzed by cellulases. As these polymers are nearly exclusively linear, debranching enzymes are not likely required.

10.6 Hydrolysis of solubilized hemicellulose

The pretreatment conditions affect the solubilization, recovery, and composition of the solubilized hemicelluloses. Depending on the raw material and pretreatment conditions, high molecular weight fractions, oligosaccharides, monomers, or sugar degradation products are formed. Longer pretreatment times (at around 190°C) and additives, such as SO₂, lead generally to better recovery of monomers. Less severe pretreatment conditions lead to solubilization of xylan and/or formation of oligosaccharides which can be hydrolyzed into monomers by enzymes (101). The total enzymatic hydrolysis of substituted oligomers needs the synergistic action of endoenzymes and accessory enzymes. The raw material, pretreatment conditions, and thus the structure of the solubilized hemicellulose oligomers should be known for the identification of enzymes needed.

The solubilized fraction from steam-pretreated birch wood contained about 10% acetyl groups, which were liberated with a culture filtrate from *T. reesei* (77). Synergy between

xylanases, β -xylosidase, and acetyl esterase of *T. reesei* was shown to be essential for the production of xylose from steamed birch xylan. Hydrolysis of the high molecular weight fraction of steamed birch wood hemicellulose by xylanase alone produced only about 10% of the amount of xylose produced by the whole set of enzymes (see Table 10.2). The work of characterizing oligomers and solubilized polymers from other substrates and pretreatment techniques is ongoing, with the details being scarce and widely varied. In essence, the best enzymes for a specific hydrolysate must be worked out empirically on a case-by-case basis. Knowledge of the structures involved, either in the native feedstock or hydrolysate, will provide significant clues; however, the lack of details regarding enzyme substrate specificities will still necessitate extensive screening and synergy studies.

Acknowledgment

This work was supported by the US DOE Office of the Biomass Program.

References

1. Curling, S.F., Clausen, C.A. & Winandy, J.E. (2002) Relationships between mechanical properties, weight loss, and chemical composition of wood during incipient brown-rot decay. *Forest Products Journal*, **52**, 34–39.
2. Kroon, P.A. & Williamson, G. (1996) Release of ferulic acid from sugar-beet pulp by using arabinanase, arabinofuranosidase and an esterase from *Aspergillus niger*. *Biotechnology and Applied Biochemistry*, **23**, 263–267.
3. Pena, M.J. & Carpita, N.C. (2004) Loss of highly branched arabinans and debranching of rhamnogalacturonan I accompany loss of firm texture and cell separation during prolonged storage of apple. *Plant Physiology*, **135**, 1305–1313.
4. Laine, C. & Tamminen, T. (2002) Origin of carbohydrates dissolved during oxygen delignification of birch and pine kraft pulp. *Nordic Pulp and Paper Research Journal*, **17**, 168–171.
5. Luonteri, E., Laine, C., Uusitalo, S., Teleman, A., Siika-aho, M. & Tenkanen, M. (2003) Purification and characterization of *Aspergillus* beta-D-galactanases acting on beta-1,4- and beta-1,3/6-linked arabinogalactans. *Carbohydrate Polymers*, **53**, 155–168.
6. Willfor, S., Sjöholm, R., Laine, C., Roslund, M., Hemming, J. & Holmbom, B. (2003) Characterisation of water-soluble galactoglucomannans from Norway spruce wood and thermomechanical pulp. *Carbohydrate Polymers*, **52**, 175–187.
7. Willfor, S., Sjöholm, R., Laine, C. & Holmbom, B. (2002) Structural features of water-soluble arabinogalactans from Norway spruce and Scots pine heartwood. *Wood Science and Technology*, **36**, 101–110.
8. Smith, M.M. & Hartley, R.D. (1983) Occurrence and nature of ferulic acid substitution of cell-wall polysaccharides in graminaceous plants. *Carbohydrate Research*, **118**, 65–80.
9. Christov, L.P. & Prior, B.A. (1993) Esterases of xylan-degrading microorganisms – production, properties, and significance. *Enzyme and Microbial Technology*, **15**, 460–475.
10. Tenkanen, M., Luonteri, E. & Teleman, A. (1996) Effect of side groups on the action of beta-xylosidase from *Trichoderma reesei* against substituted xylo-oligosaccharides. *FEBS Letters*, **399**, 303–306.
11. Faulds, C.B. & Williamson, G. (1995) Release of ferulic acid from wheat bran by a ferulic acid esterase (Fae-Iii) from *Aspergillus niger*. *Applied Microbiology and Biotechnology*, **43**, 1082–1087.

12. Tenkanen, M., Thornton, J. & Viikari, L. (1995) An acetylglucosaminase of *Aspergillus oryzae* – purification, characterization and role in the hydrolysis of O-acetyl-galactoglucosaminase. *Journal of Biotechnology*, **42**, 197–206.
13. Filho, E.X.F., Puls, J. & Coughlan, M.P. (1996) Purification and characterization of two arabinofuranosidases from solid-state cultures of the fungus *Penicillium capsulatum*. *Applied and Environmental Microbiology*, **62**, 168–173.
14. Coutinho, P.M. & Henrissat, B. (1999) Carbohydrate-active enzymes: An integrated database approach. In: *Recent Advances in Carbohydrate Bioengineering* (eds H.J. Gilbert, G. Davies, B. Henrissat & B. Svensson). The Royal Society of Chemistry, Cambridge.
15. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. & Bairoch, A. (2003) ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, **31**, 3784–3788.
16. Barthelme, J., Ebeling, C., Chang, A., Schomburg, I. & Schomburg, D. (2007) BRENDA, AMENDA and FRENDA: The enzyme information system in 2007. *Nucleic Acids Research*, **35**, D511–D514.
17. Biely, P., Vrsanska, M., Tenkanen, M. & Kluepfel, D. (1997) Endo-beta-1,4-xylanase families: Differences in catalytic properties. *Journal of Biotechnology*, **57**, 151–166.
18. Shareck, F., Roy, C., Yaguchi, M., Morosoli, R. & Kluepfel, D. (1991) Sequences of 3 genes specifying xylanases in *Streptomyces lividans*. *Gene*, **107**, 75–82.
19. Sakka, K., Yoshikawa, K., Kojima, Y., Karita, S., Ohmiya, K. & Shimada, K. (1993) Nucleotide-sequence of the *Clostridium stercorarium* xyla gene encoding a bifunctional protein with beta-D-xylosidase and alpha-L-arabinofuranosidase activities, and properties of the translated product. *Bioscience Biotechnology and Biochemistry*, **57**, 268–272.
20. Stalbrand, H., Saloheimo, A., Vehmaanpera, J., Henrissat, B. & Penttila, M. (1995) Cloning and expression of beta-mannanase from *Trichoderma reesei*. *Abstracts of Papers of the American Chemical Society*, **209**, 195-BIOT.
21. Tenkanen, M., Buchert, J. & Viikari, L. (1995) Binding of hemicellulases on isolated polysaccharide substrates. *Enzyme and Microbial Technology*, **17**, 499–505.
22. Sabini, E., Schubert, H., Murshudov, G., Wilson, K.S., Siika-Aho, M. & Penttila, M. (2000) The three-dimensional structure of a *Trichoderma reesei* beta-mannanase from glycoside hydrolase family 5. *Acta Crystallographica Section D – Biological Crystallography*, **56**, 3–13.
23. McCleary, B.V. (1991) Comparison of endolytic hydrolases that depolymerize 1,4-beta-D-mannan, 1,5-alpha-L-arabinan, and 1,4-beta-D-galactan. *ACS Symposium Series*, **460**, 437–449.
24. Kusakabe, I., Park, G.G., Kumita, N., Yasui, T. & Murakami, K. (1988) Specificity of beta-mannanase from *Penicillium purpurogenum* for konjac glucosaminase. *Agricultural and Biological Chemistry*, **52**, 519–524.
25. Tenkanen, M., Makkonen, M., Perttula, M., Viikari, L. & Teleman, A. (1997) Action of *Trichoderma reesei* mannanase on galactoglucosaminase in pine kraft pulp. *Journal of Biotechnology*, **57**, 191–204.
26. Sabini, E., Wilson, K.S., Siika-aho, M., Boisset, C. & Chanzy, H. (2000) Digestion of single crystals of mannan I by an endo-mannanase from *Trichoderma reesei*. *European Journal of Biochemistry*, **267**, 2340–2344.
27. Hagglund, P., Sabini, E., Boisset, C., Wilson, K., Chanzy, H. & Stalbrand, H. (2001) Degradation of mannan I and II crystals by fungal endo-beta-1,4-mannanases and a beta-1,4-mannosidase studied with transmission electron microscopy. *Biomacromolecules*, **2**, 694–699.
28. Margolles-Clark, E., Tenkanen, M., Soderlund, H. & Penttila, M. (1996) Acetyl xylan esterase from *Trichoderma reesei* contains an active-site serine residue and a cellulose-binding domain. *European Journal of Biochemistry*, **237**, 553–560.

29. Ademark, P., Lundqvist, J., Hagglund, P., Tenkanen, M., Torto, N., Tjerneld, F. & Stalbrand, H. (1999) Hydrolytic properties of a beta-mannosidase purified from *Aspergillus niger*. *Journal of Biotechnology*, **75**, 281–289.
30. Keitel, T., Thomsen, K.K. & Heinemann, U. (1993) Crystallization of barley (1-3,1-4)-beta-glucanase, isoenzyme-Ii. *Journal of Molecular Biology*, **232**, 1003–1004.
31. Brummell, D.A., Catala, C., Lashbrook, C.C. & Bennett, A.B. (1997) A membrane-anchored E-type endo-1,4-beta-glucanase is localized on Golgi and plasma membranes of higher plants. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 4794–4799.
32. Kotake, T., Nakagawa, N., Takeda, K. & Sakurai, N. (1997) Purification and characterization of wall-bound exo-1,3-beta-D-glucanase from barley (*Hordeum vulgare* L) seedlings. *Plant and Cell Physiology*, **38**, 194–200.
33. Hu, G.G. & Rijkenberg, F.H.J. (1998) Subcellular localization of beta-1,3-glucanase in *Puccinia recondita* f.sp. tritici-infected wheat leaves. *Planta*, **204**, 324–334.
34. Hrmova, M. & Fincher, G.B. (2001) Structure–function relationships of beta-D-glucan endo- and exohydrolases from higher plants. *Plant Molecular Biology*, **47**, 73–91.
35. Leubner-Metzger, G. & Meins, F. (2001) Antisense-transformation reveals novel roles for class I beta-1,3-glucanase in tobacco seed after-ripening and photodormancy. *Journal of Experimental Botany*, **52**, 1753–1759.
36. Ståhlberg, J. (1991) *Functional Organization of Cellulases from Trichoderma reesei*. Uppsala University, Sweden.
37. Nidetzky, B. & Claeysens, M. (1994) Specific quantification of *Trichoderma reesei* cellulases in reconstituted mixtures and its application to cellulase–cellulose binding-studies. *Biotechnology and Bioengineering*, **44**, 961–966.
38. Shoemaker, S., Watt, K., Tsitovsky, G. & Cox, R. (1983) Characterization and properties of cellulases purified from *Trichoderma reesei* strain-L27. *Bio-Technology*, **1**, 687–690.
39. Henrissat, B., Driguez, H., Viet, C. & Schulein, M. (1985) Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Bio-Technology*, **3**, 722–726.
40. Macarron, R., Acebal, C., Castillon, M.P. & Claeysens, M. (1996) Mannanase activity of endoglucanase III from *Trichoderma reesei* QM9414. *Biotechnology Letters*, **18**, 599–602.
41. Karlsson, J., Momcilovic, D., Wittgren, B., Schulein, M., Tjerneld, F. & Brinkmalm, G. (2002) Enzymatic degradation of carboxymethyl cellulose hydrolyzed by the endoglucanases Cel5A, Cel7B, and Cel45A from *Humicola insolens* and Cel7B, Cel12A and Cel45Acore from *Trichoderma reesei*. *Biopolymers*, **63**, 32–40.
42. McQueen-Mason, S.J., Fry, S.C., Durachko, D.M. & Cosgrove, D.J. (1993) The relationship between xyloglucan endotransglycosylase and in vitro-cell wall extension in cucumber hypocotyls. *Planta*, **190**, 327–331.
43. Fanutti, C., Gidley, M.J. & Reid, J.S.G. (1996) Substrate subsite recognition of the xyloglucan endotransglycosylase or xyloglucan-specific endo-(1->4)-beta-D-glucanase from the cotyledons of germinated nasturtium (*Tropaeolum majus* L) seeds. *Planta*, **200**, 221–228.
44. McQueen-Mason, S. (1997) Plant cell walls and the control of growth. *Biochemical Society Transactions*, **25**, 204–214.
45. Crombie, H.J., Chengappa, S., Hellyer, A. & Reid, J.S.G. (1998) A xyloglucan oligosaccharide-active, transglycosylating beta-D-glucosidase from the cotyledons of nasturtium (*Tropaeolum majus* L) seedlings – purification, properties and characterization of a cDNA clone. *Plant Journal*, **15**, 27–38.
46. Desveaux, D., Faik, A. & MacLachlan, G. (1998) Fucosyltransferase and the biosynthesis of storage and structural xyloglucan in developing nasturtium fruits. *Plant Physiology*, **118**, 885–894.
47. Nicol, F., His, I., Jauneau, A., Vernhettes, S., Canut, H. & Hofte, H. (1998) A plasma membrane-bound putative endo-1,4-beta-D-glucanase is required for normal wall assembly and cell elongation in *Arabidopsis*. *EMBO Journal*, **17**, 5563–5576.

48. Schroder, R., Atkinson, R.G., Langenkamper, G. & Redgwell, R.J. (1998) Biochemical and molecular characterisation of xyloglucan endotransglycosylase from ripe kiwifruit. *Planta*, **204**, 242–251.
49. Campbell, P. & Braam, J. (1999) Xyloglucan endotransglycosylases: Diversity of genes, enzymes and potential wall-modifying functions. *Trends in Plant Science*, **4**, 361–366.
50. Cosgrove, D.J. (1999) Enzymes and other agents that enhance cell wall extensibility. *Annual Review of Plant Physiology and Plant Molecular Biology*, **50**, 391–417.
51. Rose, J.K.C. & Bennett, A.B. (1999) Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls: Parallels between cell expansion and fruit ripening. *Trends in Plant Science*, **4**, 176–183.
52. Steele, N.M. & Fry, S.C. (1999) Purification of xyloglucan endotransglycosylases (XETs): A generally applicable and simple method based on reversible formation of an enzyme-substrate complex (vol 340, pg 207, 1999). *Biochemical Journal*, **341**, 862–862.
53. Takano, M., Fujii, N., Higashitani, A., Nishitani, K., Hirasawa, T. & Takahashi, H. (1999) Endoxyloglucan transferase cDNA isolated from pea roots and its fluctuating expression in hydrotropically responding roots. *Plant and Cell Physiology*, **40**, 135–142.
54. Keegstra, K. & Raikhel, N. (2001) Plant glycosyltransferases. *Current Opinion in Plant Biology*, **4**, 219–224.
55. Sampedro, J., Sieiro, C., Revilla, G., Gonzalez-Villa, T. & Zarra, I. (2001) Cloning and expression pattern of a gene encoding an alpha-xylosidase active against xyloglucan oligosaccharides from *Arabidopsis*. *Plant Physiology*, **126**, 910–920.
56. Sarria, R., Wagner, T.A., O'Neill, M.A., Faik, A., Wilkerson, C.G., Keegstra, K. & Raikhel, N.V. (2001) Characterization of a family of *Arabidopsis* genes related to xyloglucan fucosyltransferase 1. *Plant Physiology*, **127**, 1595–1606.
57. Steele, N.M., Sulova, Z., Campbell, P., Braam, J., Farkas, V. & Fry, S.C. (2001) Ten isoenzymes of xyloglucan endotransglycosylase from plant cell walls select and cleave the donor substrate stochastically. *Biochemical Journal*, **355**, 671–679.
58. Thompson, J.E. & Fry, S.C. (2001) Restructuring of wall-bound xyloglucan by transglycosylation in living plant cells. *Plant Journal*, **26**, 23–34.
59. Kaku, T., Tabuchi, A., Wakabayashi, K., Kamisaka, S. & Hoson, T. (2002) Action of xyloglucan hydrolase within the native cell wall architecture and its effect on cell wall extensibility in azuki bean epicotyls. *Plant and Cell Physiology*, **43**, 21–26.
60. Rose, J.K.C., Braam, J., Fry, S.C. & Nishitani, K. (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature. *Plant and Cell Physiology*, **43**, 1421–1435.
61. Vincken, J.P., Beldman, G. & Voragen, A.G.J. (1997) Substrate specificity of endoglucanases: What determines xyloglucanase activity? *Carbohydrate Research*, **298**, 299–310.
62. Grishutin, S.G., Gusakov, A.V., Markov, A.V., Ustinov, B.B., Semenova, M.V. & Sinitsyn, A.P. (2004) Specific xyloglucanases as a new class of polysaccharide-degrading enzymes. *Biochimica et Biophysica Acta – General Subjects*, **1674**, 268–281.
63. Hasper, A.A., Dekkers, E., van Mil, M., van de Vondervoort, P.J.I. & de Graaff, L.H. (2002) EglC, a new endoglucanase from *Aspergillus niger* with major activity towards xyloglucan. *Applied and Environmental Microbiology*, **68**, 1556–1560.
64. Siika-aho, M., Tenkanen, M., Buchert, J., Puls, J. & Viikari, L. (1994) An alpha-glucuronidase from *Trichoderma reesei* rut C-30. *Enzyme and Microbial Technology*, **16**, 813–819.
65. Puls, J. (1997) Chemistry and biochemistry of hemicelluloses: Relationship between hemicellulose structure and enzymes required for hydrolysis. *Macromolecular Symposia*, **120**, 183–196.
66. Nagy, T., Emami, K., Fontes, C.M.G.A., Ferreira, L.M.A., Humphry, D.R. & Gilbert, H.J. (2002) The membrane-bound alpha-glucuronidase from *Pseudomonas cellulosa* hydrolyzes 4-O-methyl-D-glucuronoxyloligosaccharides but not 4-O-methyl-D-glucuronoxylan. *Journal of Bacteriology*, **184**, 4925–4929.

67. de Vries, R.P., Kester, H.C.M., Poulsen, C.H., Benen, J.A.E. & Visser, J. (2000) Synergy between enzymes from *Aspergillus* involved in the degradation of plant cell wall polysaccharides. *Carbohydrate Research*, **327**, 401–410.
68. Hata, K., Tanaka, M., Tsumuraya, Y. & Hashimoto, Y. (1992) Alpha-L-arabinofuranosidase from radish (*Raphanus sativus* L) seeds. *Plant Physiology*, **100**, 388–396.
69. Saha, B.C. (2000) alpha-L-arabinofuranosidases: Biochemistry, molecular biology and application in biotechnology. *Biotechnology Advances*, **18**, 403–423.
70. Bachmann, S.L. & McCarthy, A.J. (1991) Purification and cooperative activity of enzymes constituting the xylan-degrading system of *Thermomonospora fusca*. *Applied and Environmental Microbiology*, **57**, 2121–2130.
71. Ademark, P., Varga, A., Medve, J., Harjunpaa, V., Drakenberg, T., Tjerneld, F. & Stalbrand, H. (1998) Softwood hemicellulose-degrading enzymes from *Aspergillus niger*: Purification and properties of a beta-mannanase. *Journal of Biotechnology*, **63**, 199–210.
72. de Mancilha, I.M. & Karim, M.N. (2003) Evaluation of ion exchange resins for removal of inhibitory compounds from corn stover hydrolyzate for xylitol fermentation. *Biotechnology Progress*, **19**, 1837–1841.
73. Martin, C. & Jonsson, L.J. (2003) Comparison of the resistance of industrial and laboratory strains of *Saccharomyces* and *Zygosaccharomyces* to lignocellulose-derived fermentation inhibitors. *Enzyme and Microbial Technology*, **32**, 386–395.
74. Lima, L.H.A., Felipe, M.D.D., Vitolo, M. & Torres, F.A.G. (2004) Effect of acetic acid present in bagasse hydrolysate on the activities of xylose reductase and xylitol dehydrogenase in *Candida guilliermondii*. *Applied Microbiology and Biotechnology*, **65**, 734–738.
75. Mussatto, S.I. & Roberto, I.C. (2004) Optimal experimental condition for hemicellulosic hydrolyzate treatment with activated charcoal for xylitol production. *Biotechnology Progress*, **20**, 134–139.
76. Tenkanen, M. (1998) Action of *Trichoderma reesei* and *Aspergillus oryzae* esterases in the deacetylation of hemicelluloses. *Biotechnology and Applied Biochemistry*, **27**, 19–24.
77. Poutanen, K. & Sundberg, M. (1988) An acetyl esterase of *Trichoderma reesei* and its role in the hydrolysis of acetyl xylans. *Applied Microbiology and Biotechnology*, **28**, 419–424.
78. Poutanen, K., Sundberg, M., Korte, H. & Puls, J. (1990) Deacetylation of xylans by acetyl esterases of *Trichoderma reesei*. *Applied Microbiology and Biotechnology*, **33**, 506–510.
79. Bartolome, B., Faulds, C.B., Kroon, P.A., Waldron, K., Gilbert, H.J., Hazlewood, G. & Williamson, G. (1997) An *Aspergillus niger* esterase (ferulic acid esterase III) and a recombinant *Pseudomonas fluorescens* subsp. *cellulosa* esterase (XylD) release a 5-5' ferulic dehydrodimer (diferulic acid) from barley and wheat cell walls. *Applied and Environmental Microbiology*, **63**, 208–212.
80. Mathew, S. & Abraham, T.E. (2004) Ferulic acid: An antioxidant found naturally in plant cell walls and feruloyl esterases involved in its release and their applications. *Critical Reviews in Biotechnology*, **24**, 59–83.
81. Donaghy, J. & McKay, A.M. (1997) Purification and characterization of a feruloyl esterase from the fungus *Penicillium expansum*. *Journal of Applied Microbiology*, **83**, 718–726.
82. De Vries, R.P. & Visser, J. (1999) Regulation of the feruloyl esterase (faeA) gene from *Aspergillus niger*. *Applied and Environmental Microbiology*, **65**, 5500–5503.
83. Adams, E.L., Kroon, P.A., Williamson, G. & Morris, V.J. (2005) AFM studies of water-soluble wheat arabinoxylans – effects of esterase treatment. *Carbohydrate Research*, **340**, 1841–1845.
84. de Vries, R.P. & Visser, J. (2001) *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides. *Microbiology and Molecular Biology Reviews*, **65**, 497–522.
85. Yu, P.Q., McKinnon, J.J., Maenz, D.D., Olkowski, A.A., Racz, V.J. & Christensen, D.A. (2003) Enzymic release of reducing sugars from oat hulls by cellulase, as influenced by *Aspergillus* ferulic acid esterase and *Trichoderma* xylanase. *Journal of Agricultural and Food Chemistry*, **51**, 218–223.

86. Levasseur, A., Navarro, D., Punt, P.J., Belaich, J.-P., Asther, M. & Record, E. (2005) Construction of engineered bifunctional enzymes and their overproduction in *Aspergillus niger* for improved enzymatic tools to degrade agricultural by-products. *Applied and Environmental Microbiology*, **71**, 8132–8140.
87. Palonen, H., Thomsen, A.B., Tenkanen, M., Schmidt, A.S. & Viikari, L. (2004) Evaluation of wet oxidation pretreatment for enzymatic hydrolysis of softwood. *Applied Biochemistry and Biotechnology*, **117**, 1–17.
88. Berlin, A., Gilkes, N., Kilburn, D., Bura, R., Markov, A., Skomarovsky, A., Okunev, O., Gusakov, A., Maximenko, V., Gregg, D., Sinitsyn, A. & Saddler, J. (2005) Evaluation of novel fungal cellulase preparations for ability to hydrolyze softwood substrates – evidence for the role of accessory enzymes. *Enzyme and Microbial Technology*, **37**, 175–184.
89. Borneman, W.S., Ljungdahl, L.G., Hartley, R.D. & Akin, D.E. (1991) Isolation and characterization of p-coumaroyl esterase from the anaerobic fungus *Neocallimastix* strain MC-2. *Applied and Environmental Microbiology*, **57**, 2337–2344.
90. Tenkanen, M. & Poutanen, K. (1992) Significance of esterases in the degradation of xylans. In: *Xylans and Xylanases* (eds. J. Visser, G. Beldman, M.A. Kusters-vanSomeren & A.G.J. Voragen). Elsevier, New York.
91. Gielkens, M.M.C., Visser, J. & deGraaff, L.H. (1997) Arabinoxylan degradation by fungi: Characterization of the arabinoxylan-arabinofuranohydrolase encoding genes from *Aspergillus niger* and *Aspergillus tubingensis*. *Current Genetics*, **31**, 22–29.
92. Cybinski, D.H., Layton, I., Lowry, J.B. & Dalrymple, B.P. (1999) An acetylxytan esterase and a xylanase expressed from genes cloned from the ruminal fungus *Neocallimastix patriciarum* act synergistically to degrade acetylated xylans. *Applied Microbiology and Biotechnology*, **52**, 221–225.
93. Fillingham, I.J., Kroon, P.A., Williamson, G., Gilbert, H.J. & Hazlewood, G.P. (1999) A modular cinnamoyl ester hydrolase from the anaerobic fungus *Piromyces equi* acts synergistically with xylanase and is part of a multiprotein cellulose-binding cellulase-hemicellulase complex. *Biochemical Journal*, **343**, 215–224.
94. Coughlan, M.P. & Hazlewood, G.P. (1993) *Hemicellulose and Hemicellulases*. Portland Press, London.
95. Nishitani, K. & Nevins, D.J. (1991) Glucuronoxylan xylanohydrolase – a unique xylanase with the requirement for appendant glucuronosyl units. *Journal of Biological Chemistry*, **266**, 6539–6543.
96. Laine, C., Tamminen, T. & Hortling, B. (2004) Carbohydrate structures in residual lignin-carbohydrate complexes of spruce and pine pulp. *Holzforschung*, **58**, 611–621.
97. Irwin, D.C., Cheng, M., Xiang, B.S., Rose, J.K.C. & Wilson, D.B. (2003) Cloning, expression and characterization of a family-74 xyloglucanase from *Thermobifida fusca*. *European Journal of Biochemistry*, **270**, 3083–3091.
98. Vincken, J.P., Beldman, G. & Voragen, A.G.J. (1994) The effect of xyloglucans on the degradation of cell-wall-embedded cellulose by the combined action of cellobiohydrolase and endoglucanases from *Trichoderma viride*. *Plant Physiology*, **104**, 99–107.
99. Martinez-Fleites, C., Guerreiro, C.I.P.D., Baumann, M.J., Taylor, E.J., Prates, J.A.M., Ferreira, L.M.A., Fontes, C.M.G.A., Brumer, H. & Davies, G.J. (2006) Crystal structures of *Clostridium thermocellum* xyloglucanase, XGH74A, reveal the structural basis for xyloglucan recognition and degradation. *Journal of Biological Chemistry*, **281**, 24922–24933.
100. Murray, P.G., Grassick, A., Laffey, C.D., Cuffe, M.M., Higgins, T., Savage, A.V., Planas, A. & Tuohy, M.G. (2001) Isolation and characterization of a thermostable endo-beta-glucanase active on 1,3-1,4-beta-D-glucans from the aerobic fungus *Talaromyces emersonii* CBS 814.70. *Enzyme and Microbial Technology*, **29**, 90–98.
101. Puls, J., Poutanen, K., Korner, H.U. & Viikari, L. (1985) Biotechnical utilization of wood carbohydrates after steaming pretreatment. *Applied Microbiology and Biotechnology*, **22**, 416–423.

Chapter 11

Aerobic Microbial Cellulase Systems

David B. Wilson

11.1 Introduction

An important step in the global carbon cycle is the degradation of cellulose, the most abundant form of fixed carbon with 10^{11} tons produced by plants each year (1). Most terrestrial cellulose is degraded by cellulolytic microorganisms, primarily fungi and bacteria, although some cellulose is recycled by fire and by photodegradation (2). Aerobic microorganisms are responsible for much of the cellulose degradation in soils, but there also are many species of cellulolytic anaerobic soil bacteria such as *Clostridium thermocellum*, *C. cellulovorans*, and *Acetivibrio cellulosolvens* (3–5). Termites and some other insects are very important in cellulose degradation, especially in tropical regions, and most cellulose degrading insects contain symbiotic cellulolytic microorganisms, even though many termites and other insects produce cellulases (6, 7). Some very cellulolytic termites utilize aerobic symbiotic cellulolytic fungi to breakdown plant material in their nests and then eat the fungi and residual plant material (8). Ruminants, such as cows, sheep, and deer, also are important cellulose degrading organisms but all of the cellulose that they utilize is degraded by symbiotic rumen microorganisms, primarily bacteria. The rumen is an extremely anaerobic environment and this area has been reviewed recently (9, 10), so that these organisms will not be discussed further here.

Microorganisms catalyze cellulose degradation by producing enzymes called cellulases, which hydrolyze the β -1-4 linkages present in cellulose. Almost all cellulytic microorganisms secrete their cellulases outside their cell wall, as bacteria and fungi are unable to transport insoluble materials, like cellulose, inside the cell. The soluble sugars produced by cellulase digestion of cellulose are transported inside the cell and metabolized. Native cellulose is very resistant to hydrolysis because it is insoluble and contains crystalline regions in which the adjacent cellulose molecules have strong interactions, such as hydrogen bonds and hydrophobic stacking. Thus, the specific activities of individual cellulases are much lower than those of most enzymes. However, in terms of catalytic enhancement, cellulases are very active enzymes, as the half-life of crystalline cellulose in water at neutral pH is estimated to be about 100 million years. It takes concentrated sulfuric acid at 125°C to hydrolyze native cellulose at a reasonable rate. When cellulases are assayed on low molecular weight soluble substrates, they show normal Michaelis-Menten kinetics and some have high specific activities, showing that they are basically similar to other enzymes. However, when cellulases are assayed on insoluble substrates, they have very different properties, as the assays are

usually nonlinear with time and with the amount of enzyme. Different studies have produced different explanations for this behavior. For the endocellulase *Thermobifida fusca* Cel6A, the nonlinearity was shown to be due to substrate heterogeneity (11).

11.2 Understanding cellulases

Most cellulase assays measure the production of reducing sugars from a high molecular weight form of cellulose, as every cleavage event produces a new reducing end. Endocellulases reduce the viscosity of carboxymethylcellulose (CMC), so another way to assay them is to measure the decrease in viscosity of CMC (12). Cellulases also can be assayed by measuring the increase in the number of cellulose particles produced by a cellulase incubated with a cellulose preparation of uniform particle size (13). This assay gave different kinetics of hydrolysis than measuring reducing end increase as it was linear with time and enzyme. It needed particles that were larger than 100 μm in diameter, as smaller particles did not show an increase in number even though they gave an increase in reducing ends (14).

There are two known cellulase mechanisms: hydrolysis with retention of the stereochemistry of the anomeric hydroxyl group, and hydrolysis with inversion of the anomeric hydroxyl group (15). One important difference between these mechanisms is that most retaining enzymes can catalyze transglycosylation as well as hydrolysis, while no known inverting enzymes catalyze transglycosylation (16). Cellulases are named by the family number associated with their catalytic domain (CD) followed by a capital letter that is assigned based on the order in which family members were discovered in a given organism, with A being used for the first (17).

Cellulases are currently the third largest industrial enzyme product worldwide, by dollar volume, because of their use in cotton processing, paper recycling, as detergent enzymes, in juice extraction, and as animal feed additives. However, cellulases will become the largest volume industrial enzyme, if ethanol, butanol, or some other fermentation product of sugars produced from biomass becomes a major transportation fuel, as seems likely. Currently, industrial cellulases are almost all produced from aerobic cellulolytic fungi, such as *Hypocrea jecorina* (*Trichoderma reesei*) or *Humicola insolens*. This is due to the ability of these organisms to produce extremely large amounts of crude cellulase (about 130 g/L), the relatively high specific activity of their crude cellulase on crystalline cellulose, and the ability to genetically modify these strains to tailor the set of enzymes they produce, so as to give optimal activity for specific uses.

Most aerobic cellulolytic microorganisms secrete a set of individual cellulases, which contain a carbohydrate-binding module (CBM) joined by a flexible linker peptide to the CD, and additional domains are often present. In some cellulases, the CBM is N-terminal to the CD, while in others it is C-terminal and the location probably does not affect its function. In contrast, most anaerobic microorganisms produce large (>1 million MW) multienzyme complexes, called cellulosomes, which are usually bound to the outer surface of the microorganism (18, 19). Only a few of the enzymes in cellulosomes contain a CBM, but the scaffoldin protein to which they are attached does contain a CBM, which binds the complex to cellulose. In both aerobic and anaerobic organisms, certain cellulases can act synergistically on crystalline cellulose with the specific activity of some mixtures being

more than ten times that of any single cellulase in the mixture (20). Even though cellulose is a homopolymer of glucose, with only a single type of linkage (β -1-4), and with the disaccharide cellobiose being the repeating unit, cellulases are very diverse in their structures, mechanisms, and sequences.

11.3 Diversity of cellulases

There are 14 cellulase families listed on the CAZy web site: (<http://afmb.cnrsmrs.fr/CAZY/fam/acc.GH.html>). Several of these families, 10, 26, 51, 74, mainly contain other types of glycosyl hydrolases, with only a few members having cellulase activity but, even if these families are excluded, there are still ten cellulase families. The enzymes in any given family show significant sequence homology with some or all of the other family members. All members of a glycosyl hydrolase family have the same basic protein fold and utilize the same catalytic mechanism but their substrate specificities can be quite different. Because there is a wide range of amino acid sequences that can give the same protein fold, several families share the same fold, even though the sets of sequences in each family show little similarity between the families.

There are nearly twice as many cellulase families, as are present in the next largest group of hydrolases, the seven xylanase families. Furthermore, there are seven different protein folds among the known cellulase structures, and a structure is not yet known for one family. There are two possible reasons why there is so much cellulase diversity. One is that the actual substrate of most cellulases is not pure cellulose but rather plant cell walls, which are extremely diverse and complex, containing many other components, some of which are bound to the cellulose fibrils (21). The other reason is that cellulose itself is quite complex with both crystalline and amorphous regions. It appears that cellulases are under positive selection, as when the DNA and protein sequences of two related cellulase genes were compared there were nearly as many DNA changes that caused an amino acid change (nonsynonymous) as there were DNA changes that did not change the amino acid (synonymous) (22).

There are three functionally different types of cellulases: endocellulases, also called endoglucanases, exocellulases, also called cellobiohydrolases, and processive endocellulases, which were discovered later (23). To completely hydrolyze cellulose to glucose, a fourth enzyme, β glucosidase, is required, which hydrolyzes the soluble oligosaccharides produced by the cellulases to glucose. Many aerobic fungi secrete a β glucosidase as part of their crude cellulase, while most cellulolytic aerobic bacteria do not, and their β glucosidases are usually cytoplasmic. Some organisms, mainly anaerobic bacteria, contain cellobiose phosphorylase, also called dextrin phosphorylase, which converts cellobiose and soluble dextrans to glucose and glucose-1-phosphate, conserving the energy in the cellobiose linkage (24). All endocellulase CDs, whose structures have been determined, have an open active site, as would be expected, since they are able to bind to the interior of long cellulose molecules (25). In contrast, all exocellulases have their active sites in a tunnel, consistent with their processive activity (26). In the case of glycosyl hydrolase family GH-48 enzymes, only part of the active site is in the tunnel, but these enzymes are just as processive as family GH-7 enzymes, where the entire active site is in the tunnel (27). There are two classes of exocellulases (28); one class attacks the nonreducing end of a cellulose molecule and all known members of this class are

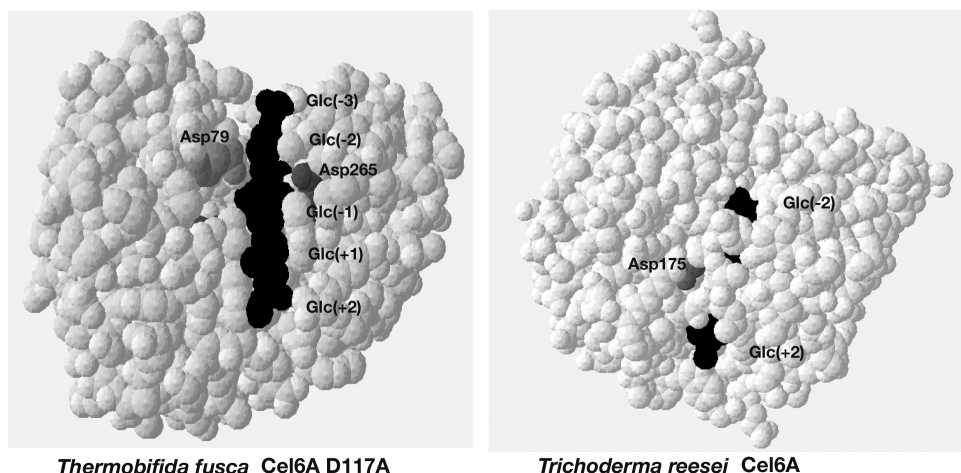


Figure 11.1 Model of the three-dimensional structures of the catalytic domains of the endocellulase, *T. fusca* Cel6A and the exocellulase, *T. reesei* Cel6A.

in family GH-6. Members of the other class attack the reducing end of a cellulose chain and all aerobic fungal members of this class are in family GH-7, while the bacterial members are in family GH-48 (see Figure 11.1). It is interesting that the anaerobic fungal members of this class are in family GH-48, rather than in family GH-7 (29). All exocellulases act processively, sequentially cleaving cellobiose residues from a cellulose molecule, so that they are also called cellobiohydrolases. It has been claimed that the *T. reesei* exocellulase, Cel6A, can act as an endocellulase and that is the reason it can synergize with *T. reesei* exocellulase, Cel7A (30); however, it has been shown that all of the hydrolysis in a synergistic mixture of these two enzymes results from exocellulolytic activity (20).

There are a number of claims in the literature that specific enzymes are exocellulases, when they are actually endocellulases. In particular, *Clostridium thermocellum*, CBHA (31) is clearly an endocellulase, as shown by the open active site seen in its X-ray structure (32) and this was confirmed by a set of assays, which showed that it behaved like an endocellulase in three different assays: higher activity on CMC than other substrates, reducing the viscosity of CMC and producing 40% insoluble reducing sugars from filter paper, while exocellulases produce from 5 to 8% insoluble reducing sugars from filter paper (33). Another example is Cel6A from the anaerobic rumen fungus, *Neocallimastix patriciarum*, which has very high activity relative to other family GH-6 exocellulases but not relative to family GH-6 endocellulases (34). By all the above tests, this enzyme turned out to be a true endocellulase (Wilson, D.B., unpublished). It is often stated that an enzyme is an exocellulase because it produces cellobiose as its major soluble product, but this is true of many endocellulases. Some workers have claimed that only exocellulases have activity on para-nitrophenyl- β -cellobioside but that is not true, as many endocellulases hydrolyze this substrate.

All well-documented processive endocellulases are in family G-9, which is the largest cellulase family and includes most plant cellulases, animal cellulases, many bacterial cellulases

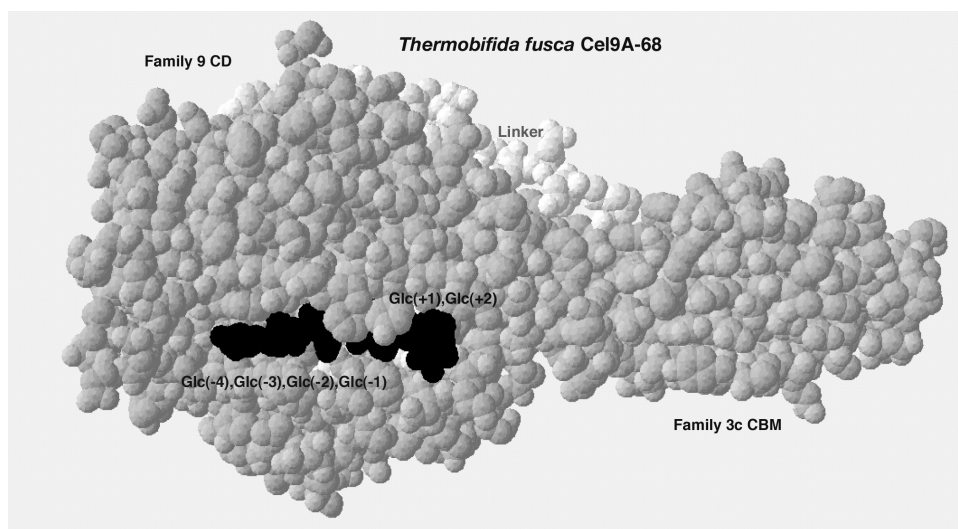


Figure 11.2 Model of the three-dimensional structure of the 68 kDa form of *T. fusca* Cel9A.

and surprisingly, very few fungal cellulases (see Figure 11.2). Processive endoglucanases have an open active site cleft like all endocellulases but in addition they contain a family 3 CBM, which is rigidly attached to the C-terminus of the CD (35). The two domains are oriented so that a cellulose chain can bind simultaneously to both domains. The family 3c CBMs, that are present in processive endocellulases, differ from families 3a and 3b CBMs, in that they lack the conserved aromatic residues, which cause the high affinity for cellulose. Although the 3c CBMs bind very weakly to cellulose, it has been shown that they are necessary for the processive activity of these enzymes (36).

There do not appear to be major differences between the CD families of the cellulases present in cellulosomes and the families of cellulases secreted by aerobic microorganisms, as most cellulase families contain cellulases from both types of microorganisms. However, all known GH-7 cellulases are from aerobic fungi or termites, and there are no known GH-6 cellulases produced by anaerobic bacteria. Furthermore, all GH-12 cellulases appear to be produced by aerobic microorganisms, but this is currently a small family that only contains endocellulases. At this time, all known GH-48 cellulases are exocellulases and this is the only cellulase family that does not contain endocellulases.

There have been a number of studies that analyzed the properties of the cellulose that remained after significant hydrolysis had occurred by a pure cellulase, to try to identify the preferred sites of attack for that cellulase; i.e., amorphous or crystalline regions, as well as how the enzyme has changed the average chain length of the cellulose. A study of *T. reesei* Cel7A, an exocellulase, and *T. reesei* Cel7B, an endocellulase, found as expected that Cel7A did not cause large changes in the cellulose chain length while Cel7B did (37). A study of four *Cellulomonas fimi* cellulases acting on Sigma cellulose found that Cel5A rapidly reduced the chain length but Cel6A had a lesser effect on chain length, even though it also is an endocellulase. Both enzymes increased the crystallinity of the residual cellulose, suggesting that they preferentially degrade amorphous regions in the cellulose. The two exocellulases

tested, Cel6B and Cel48A, had no effect on chain length and Cel6B increased crystallinity of the residual cellulose while Cel48A decreased its crystallinity (38). Four synergistic mixtures were tested and none of them caused significant differences in crystallinity. Another study of comparable enzymes from *T. fusca* showed that both the endocellulase Cel5A and the exocellulase Cel6B primarily digested amorphous cellulose, while the processive endocellulase Cel9A digested both types of cellulose (39).

11.4 Cellulose-binding domains

CBMs play an important role in the ability of cellulases to degrade crystalline cellulose. They have little or no effect on the activity of most cellulases on a soluble cellulose derivative, CMC, amorphous cellulose or oligosaccharides (40). One role of a CBM is to anchor the cellulase to the insoluble cellulose, so that the CD remains close to the substrate. The flexible linker that separates the CBM from the CD allows the CD to access regions of the cellulose, adjacent to the bound CBM. Some workers have proposed that a CBM can also disrupt the structure of cellulose, making it more accessible to the CD (41, 42) but this is still controversial (43). This activity would be equivalent to the Cx activity proposed by Reese, in his early discussion of the nature of cellulases (44). CBMs have also been reported to target cellulases to specific regions of cellulose, presumably to the regions where they will be most active (45). It has been shown that family 2 CBMs can diffuse on the surface of cellulose without dissociation; giving them the ability to readily access new regions of a cellulose particle after the CD has hydrolyzed cellulose near the original site of binding (46). Other CBMs that bind to crystalline cellulose may also have this ability, since they appear to bind in a similar way.

There are many CBM families with 45 listed on the CAZy web site (http://afmb.cnrs-mrs.fr/CAZY/fam/acc_CBM.html). Not all CBMs bind cellulose, as many families contain chitin-binding domains, xylan-binding domains, or mannose-binding domains. Some CBMs can bind to several polymers, while others are specific for only one. Labeled CBMs are being used to stain plant materials and different members of a given family can give very different staining, showing that there is even greater binding specificity than is seen with pure substrates (47). Almost all known enzymes that have high activity on insoluble substrates contain a substrate-binding domain, in addition to a CD, so that the presence of such a domain is a general property of this type of enzyme.

All fungal cellulase CBMs are in family I and they are small, containing about 30 AA. Most aerobic bacterial cellulase CBMs are in family 2 and they are larger, containing about 120 AA. The CBMs on cellulosomal scaffoldins are in family 3. Most of the CBMs in these three families bind to crystalline cellulose and have a relatively flat-binding surface that usually contains three aromatic residues spaced, so they can bind to three adjacent glucose residues in a cellulose molecule. They also contain a number of residues, which can hydrogen bond to the cellulose chain, but site-directed mutagenesis has shown that the aromatic residues are essential for high affinity binding, while the other residues play a secondary role (48). The CBMs in families 4 and 6 bind to single cellulose molecules and their binding sites are in a groove (49). A number of cellulases contain multiple CBMs and in some cases they are from the same family and in other cases they are from different families. It has been shown that the affinity of a protein containing two CBMs can be significantly higher than one with

only one CBM (50). Atomic force microscopy of the binding of a family I CBM to cellulose found that the bound CBM was present in aggregates, not as single domains (49).

Binding of a cellulase to cellulose is an important step in hydrolysis and there have been many studies of cellulase binding (50–54). Cellulases that contain a CBM bind tightly to cellulose, but cellulase CDs bind cellulose weakly. Mutations that dramatically reduce activity can lead to enhanced CD binding, showing that the cleavage products bind more weakly than the intact substrate. The extent of binding is directly related to the accessible surface area of the substrate, and accessible surface area also determines the rate of hydrolysis of a substrate. Much of the surface area of cellulose is in the interstitial region between microfibrils of cellulose (irregularly shaped pores), so that the size of a cellulase will affect how much of the cellulose surface area is available for its binding. There is good evidence that binding to cellulose does not fit the Langmuir isotherm, and in many cases binding appears to be irreversible (53). At low protein concentrations, several family 2 CBMs were shown to bind reversibly and in this region binding might be to the external surface, but at higher extents of binding, where binding might be occurring in pores, it was irreversible (53). In a study where the binding of a mixture of two cellulases was studied, synergism in binding was observed in several mixtures (54).

11.5 Cellulase synergism

Synergism is usually only seen in the digestion of substrates that contain crystalline cellulose, probably because there are only a few regions in this substrate that are accessible to each cellulase. It seems likely that synergism occurs only when two cellulases attack different regions of the cellulose molecule and each one creates new sites of attack for the other enzymes in the mixture. There is no evidence that synergism requires interactions between the synergizing cellulases, as cellulases from unrelated organisms show similar synergism to those from the same organism. All good endocellulases are able to show synergism with any exocellulase, but most endocellulases do not show synergism with each other. Exocellulases show synergism with other exocellulases, but only if they attack different ends of the cellulose chains. Processive endocellulases can give synergy with all other types of cellulase (Wilson, D.B., unpublished). It is interesting that cellulose pretreated with an endocellulase is a better substrate for exocellulases than is untreated cellulose, but the reverse is not true (55). This is true, even though it has been shown in synergistic mixtures of an exocellulase and an endocellulase that endocellulase activity is increased as much in the mixture as is exocellulase activity (28). It seems likely that when an exocellulase degrades a cellulose chain, it disrupts some regions in adjacent cellulose molecules, making them more available to an endocellulase. However, over time the molecules in the disrupted regions may reform their interactions with adjacent chains, so that when an endocellulase is added, after exocellulase treatment, it is not able to utilize the transiently disrupted chains.

11.6 Cellulases from *Trichoderma reesei*

The most studied aerobic cellulolytic microorganism is the fungus, *Hypocrea jecorina*, originally called *Trichoderma reesei*. It was isolated and studied by Drs Reese and Mandels at the

Army Quartermaster Lab in Natick, MA, during World War II, because it was degrading the cotton fabrics used by the army for tents, gun straps, etc. on islands in the Pacific Ocean (56). The original goal of this work was to find cellulase inhibitors, which was not achieved, as only the toxic ions, Hg and Ag, are good inhibitors. However, this group also carried out many studies on the organism and its crude cellulase, which then led to the development of high producing mutant strains by Dr Eveleigh that were used to develop the strains used for industrial cellulase production by several companies (57). The most abundant cellulase produced by *T. reesei* is the reducing end-specific exocellulase, Cel7A (cellobiohydrolase I), which makes up about 70% of the cellulase protein secreted by *T. reesei* (58). The next most abundant cellulase is Cel6A (CBH II), which makes up a further 10% of *T. reesei* secreted cellulase. *T. reesei* crude cellulase contains seven endoglucanases of which Cel7B (EGLI) is the most abundant. In addition, Cel5A (EGLII), Cel12A (EGLIII), Cel61A (EGLIV), Cel45A (EGLV), Cel5B, and Cel61B are also present in *T. reesei* secreted cellulase. Most of the *T. reesei* cellulases contain a family I CBM except Cel5B, Cel12A, and Cel61B. It is not clear why *T. reesei* produces so many endocellulases, but Cel12A was shown to have expansin activity as well as cellulase activity (59). Expansin is present in plants and it appears to disrupt the hydrogen bonds that bind different carbohydrate chains together in plant cell walls, so that it may make the chains more accessible to hydrolytic enzymes. The least studied endocellulase is Cel61A, which has extremely low cellulase activity. It is quite surprising that when a set of thermophilic fungal cellulases were screened for the ability to stimulate the activity of *T. reesei* crude cellulase, a number of them were able to increase it about threefold and the component that was most active in giving this stimulation was a family 61 enzyme (60). Little is known about the role of Cel45A in cellulose degradation (61). Another protein secreted by *T. reesei* is swollenin, which is a low molecular weight protein that has no catalytic activity but appears to disrupt the structure of cellulose microfibrils, possibly by breaking hydrogen bonds (62).

Most of the *T. reesei* cellulases are glycosylated and glycosylation appears to protect the cellulases from proteolysis (63). The linker peptide is particularly susceptible to proteolysis and *T. reesei* secretes proteases (64), so that protection from proteolysis may be an important role for the O-linked glycosylation found on the linker peptide (65). The role of the N-linked glycosylation on the CD is unclear at this time. There is a great deal of heterogeneity in the glycosylation of any given cellulase and this causes heterogeneity of each enzyme during gel electrophoresis and column chromatography (66).

There have been extensive studies of the regulation of cellulase synthesis in *T. reesei* and it appears that regulation is complex (67). Glucose strongly represses cellulase synthesis and the β -1,2-linked glucose disaccharide, sepharose, induces synthesis. A number of transcription factors have been identified in *T. reesei*, which can bind to cellulase promoters, and some of these are activators and some are repressors. The exact mechanisms that regulate cellulase synthesis are still not completely understood.

11.7 Other fungal cellulases

Another organism used for industrial cellulase production is *Humicola insolens*, which seems to produce the same set of cellulases as *T. reesei* except that it produces a family GH-6 endoglucanase (68). However, these two organisms are not closely related, even though they

are both brown rot fungi, which do not degrade lignin. *Phanerochaete chrysosporium* is a white rot fungus that degrades lignin, while not degrading much cellulose, despite containing a set of cellulase genes (69). A surprising finding is that it contains seven CBH I genes that are differentially regulated, in contrast to the single CBH I gene in *T. reesei*. The cellulases produced by *Aspergillus aculeatus* have been extensively studied, and it produces nine cellulases of which three have been sequenced: Cel7A, Cel12A, and Cel5A. From this limited data, it seems that its cellulases may be similar to those of *T. reesei* (70). Another fungus whose cellulases have been studied is *Talaromyces emersonii*, which produces two exocellulases: Cel7A and Cel6A, which have been extensively studied, and several endocellulases of which only Cel5A has been studied (71). *Chrysosporium lucknowense* cellulases also have been studied and seem to resemble those of *T. reesei* (72). There are many other cellulolytic fungi whose cellulases have had some research including *Agaricus bisporus*, several *Aspergillus* species, *Aureobasidium pullulans*, *Cochlibolus carbonum*, several *Fusarium* species, several *Penicillium* species, *Pleurotus ostreatus*, and *Thermoascus aurantiacus*, but the total cellulase system has not been determined for any of these organisms at this time. All these aerobic fungi produce a set of individual cellulases; however, an aerobic *Chaetomium* strain has been reported to produce a large cellulase complex (73). At this time it is not known what cellulases are in the complex and if the complex is assembled using the cohesin–dockerin binding seen in cellulosomes. This system clearly should be studied further.

11.8 Cellulolytic aerobic bacteria

Almost all aerobic cellulolytic bacteria secrete sets of individual cellulases, and most of these cellulases contain one or more CBMs. Two such organisms, whose cellulases have been well characterized, are *Cellulomonas fimi* and *T. fusca*. The sets of six cellulases produced by these two organisms are similar in their activity, CD families and CBMs, but differ significantly in their sequences and domain order, suggesting that these organisms did not obtain their cellulase genes from a common ancestor, but rather each set is the result of convergent evolution. In five of the six corresponding cellulase pairs from each species, the family 2 CBM is at the opposite end of the enzyme (23). Both these organisms are actinomycetes and they are found in soils, but *C. fimi* is mesophilic with an optimum growth temperature near 30°C, while *T. fusca* is moderately thermophilic with an optimum growth temperature of 50°C. An interdisciplinary group at the University of British Columbia has studied *C. fimi* cellulases, and this group has made many important contributions to cellulase research (74).

T. fusca is often found in compost piles, rotting hay, and manure piles. The genome sequence of *T. fusca* was determined in 2000 by the DOE Joint Genome Institute and the finished 3.7 mb sequence is available at: http://genome.jgi-psf.org/finished_microbes/thefu/thefu.home.html (75). These two organisms are not closely related, as *C. fimi* is in the suborder *Micrococcineae* while *T. fusca* is in the suborder *Streptosporangineae*. It is interesting that *Streptomyces lividans*, which is a relative of *T. fusca*, contains five cellulase genes that are similar to those present in *T. fusca* and, in every one, the family 2 CBM is in the same location as it is in *T. fusca*, suggesting that these sets of genes did come from a common ancestor.

In addition to the six cellulases, *T. fusca* secretes a number of other proteins when it is grown on cellulose. Among them are a xyloglucanase, a xylanase, a family 81 β -1,3 endoglucanase, and two proteins that bind to cellulose but do not appear to have any catalytic activity (76–79). The xyloglucanase and the xylanase both contain a family 2 CBM and bind tightly to cellulose. The CBM on the xylanase also binds to xylan, which is unusual, as most family 2 CBMs bind only to one of these polymers (80). It seems likely that the role of the xyloglucanase is to degrade xyloglucan, which is bound to cellulose, allowing the cellulases to access the cellulose rather than to allow *T. fusca* to utilize xyloglucan, since *T. fusca* does not grow on xyloglucan, even though it completely degrades it to oligomers (76). Further evidence for this conclusion is that a xyloglucan–cellulose composite, which was not degraded by a mixture of pure cellulases, was degraded when pure *T. fusca* xyloglucanase was added to the cellulase mixture (76). The role of the two cellulose-binding proteins is not known, but they do stimulate the activity of several *T. fusca* cellulases in the initial phase of the reaction, when the cellulases are assayed at low levels (79).

Cellulase synthesis in *T. fusca* and many cellulolytic bacteria is regulated by at least two mechanisms: induction by cellobiose and laminaribiose (β -1,3-glucose disaccharide), and repression by any good carbon source including cellobiose (78, 80, 81). This makes sense, as the extracellular enzymes secreted by *T. fusca* grown on cellulose make up about 50% of the total protein in the culture. If there is sufficient sugar for growth or no cellulose, there is no reason to synthesize cellulases. The cellobiose required for induction is produced from cellulose by the uninduced level of secreted cellulase, which is mainly Cel6A. CelR is a regulatory protein that is a member of the Lac I gene family (82), and it binds to a 14 base inverted repeat sequence: TGGGAGCGCTCCCA. This sequence is upstream of the start site of all six *T. fusca* cellulase genes coding for secreted cellulases, as well as the genes for a number of other secreted proteins induced by growth on cellulose, and an operon that codes for a cytoplasmic β glucosidase and a putative cellobiose transport system (83). The CelR gene is just upstream of this operon. The binding of CelR to the regulatory sequence is inhibited by cellobiose, as expected. Since laminaribiose also induces cellulase synthesis it also probably binds to CelR but this has not been tested. At this time it is not known whether induction by laminaribiose is useful for *T. fusca*, or if it is an accidental result of the low specificity of the CelR sugar-binding site. A puzzling finding is that the Cel6B and Cel48A genes, which each encode an exocellulase, have a second CelR-binding site that is about 200 bases upstream of the start site of the gene and this site is not present in the other cellulase gene upstream sequences. These two cellulases are made in equal amounts and together they make up over 70% of the secreted cellulase protein. At this time there is no information about the mechanism by which good carbon sources inhibit cellulase synthesis in *T. fusca*.

Saccharophagus degradans is an aerobic marine plant cell wall degrading organism, whose genome was recently sequenced. Preliminary analysis of its 13 cellulase genes shows that it contains ten family GH-5 enzymes, which are most similar to endoglucanases and most of them also contain cellulose-binding domains. It also contains two family GH-9 endoglucanase genes, one of which encodes both a family 10 and a family 2 CBM (84).

Cytophaga hutchinsonii is an aerobic cellulolytic bacterium and the DOE Joint Genome Institute determined the DNA sequence of its genome (http://genome.jgi-psf.org/finished_microbes/cythu/cythu.home.html) (85). While it codes for a number of cellulase genes, most of these genes lack a CBM and all the genes appear to code for endoglucanases. These results clearly distinguish *C. hutchinsonii* from all other aerobic

cellulolytic bacteria, whose cellulases are known, and suggest that it uses a different mechanism for degrading cellulose than the secretion of a synergistic set of free cellulases used by the other well-studied aerobic cellulolytic microorganisms. None of its cellulase genes encode dockerin sequences, so that it does not appear to produce cellulosomes, like anaerobic cellulolytic bacteria. Thus, it appears that *C. hutchinsonii* has a third mechanism for degrading cellulose. It is interesting that the anaerobic rumen bacterium, *Fibrobacter succinogenes*, whose genome sequence was determined by TIGR funded by a USDA grant to the North American Consortium for Genomics of Fibrolytic Ruminant Bacteria, also does not code for any known processive cellulases and only one of the many endocellulases that have been cloned and sequenced from it appears to bind to cellulose (86, 87). *F. succinogenes* does not appear to encode dockerin domains and a scaffoldin gene has not been identified. Thus, *F. succinogenes* also probably uses a novel mechanism for degrading cellulose. *F. succinogenes* grows very rapidly on cellulose, so that its cellulose degrading mechanism is very efficient (88). One possible mechanism for these organisms is the one proposed for starch degradation by *Bacteroides thetaiotaomicron* (89). In this mechanism, starch is bound to a complex present in the outer membrane and individual molecules are transported into the periplasmic space, where they are degraded by starch degrading enzymes. This mechanism would not require processive cellulases, as individual cellulose molecules would be readily degraded by endoglucanases. If this is the process by which cellulose is degraded, it will be very interesting to determine the mechanism by which the outer membrane proteins are able to bind and transport individual cellulose molecules. It is possible that this information would allow the design of new cellulases or cellulose modifying proteins, which would be able to increase the rate of cellulose degradation by free cellulases.

Site-directed mutagenesis of cellulase genes has been used to identify residues required for activity and substrate binding. Extensive studies of *T. fusca* Cel6A have been carried out (90–92) and a number of Cel6A residues have been identified that are essential for activity or binding. From these studies, a detailed mechanism has been proposed for Cel6A, which differs from the standard cellulase mechanism in that there is no evidence for an essential catalytic base, although there is a conserved Asp residue (Asp 79), which is important for activity and it could be functioning as a nonessential catalytic base (91). Because it is in a flexible loop, it is not clear where this residue is positioned during bond cleavage, but it is unlikely to be in the position seen for the catalytic base in other cellulase families, which is on the other side of the sessile bond from the essential catalytic acid, which is Asp 117 in Cel6A (91). Loop movement is important in catalysis by Cel6A, as mutation of either of two adjacent Gly residues to Ala, which are at one end of the loop, reduce activity on CMC to 6% or 16% of WT activity (90).

In all WT Cel6A structures that contain bound sugars, the glucose molecule in subsite -1 is distorted, as is seen with most cellulases. Tyr73 is a conserved residue in family GH-6 and, when it is mutated to Phe, Cel6A activity is reduced to 10% of WT but, when it is mutated to Ser, activity is abolished (93). In a structure of the Cel6A Tyr73Ser mutant enzyme, the glucose in the -1 is not distorted (94). This result is consistent with molecular modeling of Cel6A, showing that the normal glucose conformation would overlap the position of Tyr73 (95). It also suggests that distortion of the glucose is essential for activity, since the Ser mutation eliminates both distortion and activity, while the Phe mutation, which does not eliminate the distortion only partially reduces activity (93). Tyr73 was reported to be the essential hydrophobic platform residue in family GH-6, and the mutagenesis data are

consistent with this proposal (96). Cel6A hydrolyzes CMC and amorphous cellulose (SC) 100 times faster than filter paper and 50 times faster than bacterial cellulose. Furthermore, mutations that alter key active site residues cause much greater reductions in activity on CMC and SC than on the crystalline substrates (91). These results show that there are different rate-limiting steps for the two types of substrates. It seems likely that the rate-limiting step for CMC and SC is bond cleavage, while for crystalline substrates it may be the binding of the substrate into the active site. In order to design new cellulases with higher activity on crystalline substrates, it is necessary to increase the rate of the rate-limiting step. Unfortunately, at this time the detailed mechanism by which a segment of a cellulose molecule is removed from its neighbors and bound into the active site is not known for any cellulase, so that it is not possible to rationally design mutant enzymes to increase the rate of this step.

Similar, but less extensive, studies have been carried out on the exocellulase, *T. fusca* Cel6B, and they showed that inserting a disulfide bond, which joined the two loops that cover the Cel6B active site to form a tunnel, reduced activity only 50%, so clearly loop opening, if it occurs, is not required for activity (97). These studies also found two mutations, which increased activity on crystalline cellulose, but only when the enzyme was assayed alone, not in synergistic mixtures. Several mutations were found, which reduced inhibition by cellobiose, with only a small loss of activity on crystalline cellulose. Finally, two mutations reduced activity on bacterial cellulose without reducing activity on CMC or SC, suggesting that these residues participated in a step that was only required for crystalline cellulose hydrolysis (97).

Extensive site-directed mutagenesis studies have been carried out on the processive endocellulase, *T. fusca* Cel9A (98, 99). These studies identified key residues that are important for activity, including a Glu residue that functions as the catalytic acid, and two conserved Asp residues, which are hydrogen bonded to the catalytic water molecule in the structure of the enzyme lacking bound sugars. Mutation of either residue drastically reduces activity, even though only one functions as the catalytic base (99). In the Cel9A structure, the catalytic base forms a hydrogen-bonded network with a conserved His residue and a conserved Tyr residue. Mutation of either residue drastically reduces activity, showing that this network is important for catalysis. The conserved Tyr residue is probably the hydrophobic platform residue in Cel9A, even though a different Tyr was proposed to have this function (96). When the Y429, proposed as the platform residue, was mutated, the mutant enzyme still retained about 10–30% of WT activity on various substrates, so it clearly does not have that role in catalysis (99). Several mutations increase activity on CMC, while reducing activity on bacterial cellulose. In most cases, these mutations increase the size of the active site cleft, which may allow binding of modified sugars. A number of CD mutations reduced processivity and all these mutations were in the –2 to –4 subsites and would be expected to weaken binding to these sites, suggesting that the loss of processivity was due to decreased affinity of these subsites for the cellulose chain bound to the family 3 CBM, which is part of the active site of this enzyme. This result is consistent with the results of a docking study of Cel7A, which calculated the binding energy of each of the glucose-binding subsites in this enzyme for glucose and found that the energies increased in a way to draw the cellulose chain into the active site consistent with its processive activity (100).

There have been extensive mechanistic and structural studies of GH-5 endocellulases, which are retaining hydrolases (101, 102). The GH-5 enzymes are extremely diverse in their

sequences, in their detailed structures, and in their substrate specificities. A structural study of *Acidothermus cellulolyticus* endocellulase, Cel5A (E1), showed that the eight totally conserved residues in this family were all in the active site and were close to the bound substrate (101). There are other residues, which are not conserved in this family, that also interact with the substrate, providing further evidence of the diversity in this family. Structural studies of *Bacillus agaradhaerens* endocellulase Cel5A have determined structures for every step during bond cleavage and have identified the hydrogen-bonding network to the substrate at each step (102, 103). As was found for E1, there are both conserved and nonconserved residues present in these networks. Site-directed mutagenesis of *Pyrococcus horikoshii* endocellulase Cel5A showed that the conserved nucleophile was essential for activity, but that the conserved acid/base residue was not important for activity (104). This is a very unusual finding but, since there are several other conserved residues adjacent to the acid/base residue, this result is not due to an incorrect alignment, so that it provides even more evidence for the extensive diversity of GH-5 enzymes. All of the studies described here used low molecular weight substrates and studied events during catalysis, but not the placement of a cellulose molecule into the active site, which appears to be a key step for crystalline cellulose hydrolysis.

11.9 Outlook

In the last decade, there has been a large increase in our understanding of cellulase structure functional relationships but there is still a considerable amount that needs to be learned. This is particularly true for crystalline cellulose hydrolysis, where the exact role of CBMs in hydrolysis, how cellulose molecules are bound into cellulase active sites, and how cellulose structure influences this process, need to be determined. Only when these processes are understood, will it be possible to engineer cellulase active sites to achieve more efficient hydrolysis of specific biomass substrates. This ability should improve the economics of converting biomass to liquid fuels, possibly leading to sustainable production of non-greenhouse gas producing liquid fuels.

References

1. Malhi, Y. (2002) Carbon in the atmosphere and terrestrial biosphere in the 21st century. *Philosophical Transactions, Series A, Mathematical and Physical Engineering Science*, **360**, 2925–2945.
2. Falkowski, P., Scholes, R.J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F.T., Moore, B., III, Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V. & Steffen, W. (2000) The global carbon cycle: A test of our knowledge of earth as a system. *Science*, **290**, 291–296.
3. Demain, A.L., Newcomb, M. & Wu, J.H. (2005) Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews*, **69**, 124–54.
4. Doi, R.H. & Tamaru, Y. (2001) The *Clostridium cellulovorans* cellulosome: An enzyme complex with plant cell wall degrading activity. *Chemical Record*, **1**, 24–32.
5. Lin, C., Urbance, J.W. & Stahl, D.A. (1994) *Acetivibrio cellulolyticus* and *Bacteroides cellulosolvens* are members of the greater clostridial assemblage. *FEMS Microbiology Letters*, **124**, 151–155.
6. Watanabe, H. & Tokuda, G. (2001) Animal cellulases. *Cellular and Molecular Life Sciences*, **58**, 1167–1178.

7. Ohkuma, M. (2003) Termite symbiotic systems: Efficient bio-recycling of lignocellulose. *Applied Microbiology and Biotechnology*, **61**, 1–9.
8. Shinzato, N., Muramatsu, M., Watanabe, Y. & Matsui, T. (2005) Termite-regulated fungal mono-culture in fungus combs of a macrotermite termite *Odontotermes formosanus*. *Zoological Science*, **22**, 917–922.
9. Lynd, L.R., Weimer, P.J., van Zyl, W.H. & Pretorius, I.S. (2002) Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, **66**, 506–577.
10. Desvaux, M. (2006) Unraveling carbon metabolism in anaerobic cellulolytic bacteria. *Biotechnology Progress*, **22**, 1229–1238.
11. Zhang, S., Wolfgang, D. & Wilson, D.B. (1999) Substrate heterogeneity causes the non-linear kinetics of insoluble cellulose hydrolysis. *Biotechnology and Bioengineering*, **66**, 35–41.
12. Sharrock, K.R. (1988) Cellulase assay methods: A review. *Journal of Biochemical and Biophysical Methods*, **17**, 81–105.
13. Walker, L.P., Wilson, D.B., Irwin, D.C., McQuire, C. & Price, M. (1992) Fragmentation of cellulose by the major *Thermomonospora fusca* cellulases, *Trichoderma reesei* CBHI, and their mixtures. *Biotechnology and Bioengineering*, **40**, 1019–1026.
14. Peters, L.E., Walker, L.P., Wilson, D.B. & Irwin, D.C. (1991) The impact of initial particle size on the fragmentation of cellulose by the cellulases of *Thermomonospora fusca*. *Bioresource Technology*, **35**, 313–320.
15. McCarter, J.D. & Withers, S.G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Current Opinion in Structural Biology*, **4**, 885–892.
16. Blanchard, J.E. & Withers, S.G. (2001) Rapid screening of the aglycone specificity of glycosidases: Applications to enzymatic synthesis of oligosaccharides. *Chemical Biology*, **8**, 627–633.
17. Henrissat, B., Teeri, T.T. & Warren, R.A.J. (1998) A scheme for designating enzymes that hydrolyze the polysaccharides in the cell walls of plants. *FEBS Letters*, **425**, 352–354.
18. Bayer, E.A., Belaich, J.P., Shoham, Y. & Lamed, R. (2004) The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology*, **58**, 521–554.
19. Doi, R.H. & Kosugi, A. (2004) Cellulosomes: Plant-cell-wall-degrading enzyme complexes. *Nature Reviews Microbiology*, **2**, 541–551.
20. Irwin, D.C., Spezio, M., Walker, L.P. & Wilson, D.B. (1993) Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. *Biotechnology and Bioengineering*, **42**, 1002–1013.
21. Rose, J.K.C. (ed.) (2003) *The Plant Cell Wall. Vol. 8: Annual Plant Reviews*. Blackwell Publishing, Oxford, England.
22. Lao, G., Ghangas, G.S., Jung, E.D. & Wilson, D.B. (1991) DNA sequence of three β -1,4-endoglucanase genes from *Thermomonospora fusca*. *Journal of Bacteriology*, **173**, 3397–3407.
23. Wilson, D.B. & Irwin, D. (1999) Genetics and properties of cellulases. In: *Advances in Biochemical Engineering/Biotechnology: Recent Progress in Bioconversion of Lignocellulosics*, Vol. 65 (eds. G.T. Tsao & T. Scheper), pp. 1–21. Springer-Verlag, Berlin, Germany.
24. Kim, Y.K., Kitaoka, M., Krishnareddy, M., Mori, Y. & Hayashi, K. (2002) Kinetic studies of a recombinant cellobiose phosphorylase (CBP of the *Clostridium thermocellum*) YM4 strain expressed in *Escherichia coli*. *Journal of Biochemistry (Tokyo)*, **132**, 197–203.
25. Juy, M., Amit, A.G., Alzari, P.M., Poljak, R.J., Claeysens, M., Béguin, P. & Aubert, J.-P. (1992) Three-dimensional structure of a thermostable bacterial cellulase. *Nature*, **357**, 89–91.
26. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K. & Jones, T.A. (1990) Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science*, **249**, 380–386.
27. Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J.P., Driguez, H. & Haser, R. (1998) The crystal structure of the processive endocellulase CelF of *Clostridium cellulolyticum* in complex with a thiooligosaccharide inhibitor at 2.0 Å resolution. *EMBO Journal*, **17**, 5551–5562.

28. Barr, B.K., Hsieh, Y.L., Ganem, B. & Wilson, D.B. (1996) Identification of two functionally different classes of exocellulases. *Biochemistry*, **35**, 586–592.
29. Teunissen, M.J. & Op den Camp, H.J. (1993) Anaerobic fungi and their cellulolytic and xylanolytic enzymes. *Antonie Van Leeuwenhoek*, **63**, 63–76.
30. Boisset, C., Frasnichini, C., Schulein, M., Henrissat, B. & Chanzy, H. (2000) Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. *Applied and Environmental Microbiology*, **66**, 1444–1452.
31. Zverlov, V.V., Velikodvorskaya, G.V., Schwarz, W.H., Bronnenmeier, K., Kellermann, J. & Staudenbauer, W.L. (1998) Multidomain structure and cellulosomal localization of the *Clostridium thermocellum* cellobiohydrolase CbhA. *Journal of Bacteriology*, **180**, 3091–3099.
32. Schubot, F.D., Kataeva, I.A., Chang, J., Shah, A.K., Ljungdahl, L.G., Rose, J.P. & Wang, B.C. (2004) Structural basis for the exocellulase activity of the cellobiohydrolase CbhA from *Clostridium thermocellum*. *Biochemistry*, **43**, 1163–1170.
33. McGrath, C.E. (2007) *Mechanistic and Functional Characterization of Glycosyl Hydrolases Involved in Biomass Degradation: Thermobifida fusca LAM81A and CHI18A and Clostridium thermocellum CBHA*. PhD thesis, Cornell University, Ithaca, NY.
34. Denman, S., Xue, G.P. & Patel, B. (1996) Characterization of a *Neocallimastix patriciarum* cellulase cDNA (CelA) homologous to *Trichoderma reesei* cellobiohydrolase II. *Applied and Environmental Microbiology*, **62**, 1889–1896.
35. Sakon, J., Irwin, D., Wilson, D.B. & Karplus, P.A. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Structural Biology*, **4**, 810–818.
36. Irwin, D., Shin, D.-H., Zhang, S., Barr, B.K., Sakon, J., Karplus, P.A. & Wilson, D.B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *Journal of Bacteriology*, **180**, 1709–1714.
37. Srisodsuk, M., Kleman-Leyer, K., Keranen, S., Kirk, T.K. & Teeri, T.T. (1998) Modes of action on cotton and bacterial cellulose of a homologous endoglucanase-exoglucanase pair from *Trichoderma reesei*. *European Journal of Biochemistry*, **251**, 885–892.
38. Mansfield, S.D. & Meder, R. (2003) Cellulose hydrolysis: The role of monocomponent cellulases in crystalline cellulose degradation. *Cellulose*, **10**, 159–169.
39. Chen, Y., Stipanovic, A.J., Winter, W.T., Wilson, D.B. & Kim, Y.-J. (2007) Effect of digestion by pure cellulases on crystallinity and average chain length for bacterial and microcrystalline celluloses. *Cellulose*, **14**, 283–293.
40. Shoseyov, O., Shani, Z. & Levy, I. (2006) Carbohydrate binding modules: Biochemical properties and novel applications. *Microbiology and Molecular Biology Reviews*, **70**, 283–295.
41. Din, N., Gilkes, N.R., Tekant, B., Miller, R.C., Jr., Warren, R.A.J. & Kilburn, D.G. (1991) Non-hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. *Bio/Technology*, **9**, 1096–1099.
42. Esteghlalian, A.R., Srivastava, V., Gilkes, N.R., Kilburn, D.G., Warren, R.A. & Saddle, J.N. (2001) Do cellulose binding domains increase substrate accessibility? *Applied Biochemistry and Biotechnology*, **91–93**, 575–592.
43. Bolam, D.N., Ciruela, A. & McQueen-Mason, S., Simpson, P., Williamson, M.P., Rixon, J.E., Boraston, A., Hazelwood, J.P. & Gilbert, H.J. (1998) *Pseudomonas* cellulose binding domains mediate their effects by increasing enzyme substrate proximity. *Biochemistry Journal*, **331**, 775–781.
44. Reese, E.T., Sui, R.G.H. & Levinson, H.S. (1950) The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *Journal of Bacteriology*, **59**, 485–497.
45. Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. (2004) Carbohydrate binding modules: Fine-tuning polysaccharide recognition. *Biochemistry Journal*, **382**, 769–781.

46. Jervis, E.J., Haynes, C.A. & Kilburn, D.G. (1997) Surface diffusion of cellulases and their isolated binding domains on cellulose. *Journal of Biological Chemistry*, **272**, 24016–24023.
47. McCartney, L., Blake, A.W., Flint, J., Bolam, D.N., Boraston, A.B., Gilbert, H.J. & Knox, J.P. (2006) Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules. *Proceedings of the National Academy of Sciences USA*, **103**, 4765–4770.
48. Din, N., Forsythe, I.J., Burtneck, L.D., Gilkes, N.R., Miller, R.C., Jr., Warren, R.A. & Kilburn, D.G. (1994) The cellulose-binding domain of endoglucanase A (CenA) from *Cellulomonas fimi*: Evidence for the involvement of tryptophan residues in binding. *Molecular Microbiology*, **11**, 747–755.
49. Kormos, J., Johnson, P.E., Brun, E., Tomme, P., McIntosh, L.P., Haynes, C.A. & Kilburn, D.G. (2000) Binding site analysis of cellulose binding domain CBD (N1) from endoglucanase C of *Cellulomonas fimi* by site-directed mutagenesis. *Biochemistry*, **39**, 8844–8852.
50. Linder, M., Salovuori, I., Ruohonen, L. & Teeri, T.T. (1996) Characterization of a double cellulose-binding domain. Synergistic high affinity binding to crystalline cellulose. *Journal of Biological Chemistry*, **271**, 21268–21272.
51. Nigmatullin, R., Lovitt, R., Wright, C., Linder, M., Nakari-Setälä, T. & Gama, M. (2004) Atomic force microscopy study of cellulose surface interaction controlled by cellulose binding domains. *Colloids and Surfaces. B, Biointerfaces*, **35**, 125–135.
52. Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M. & Teeri, T.T. (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proceedings of the National Academy of Sciences USA*, **100**, 484–489.
53. Jung, H., Wilson, D.B. & Walker, L.P. (2002) Binding mechanisms for *Thermobifida fusca* Cel5A, Cel6B, and Cel48A cellulose-binding modules on bacterial microcrystalline cellulose. *Biotechnology and Bioengineering*, **80**, 380–392.
54. Jeoh, T., Wilson, D.B. & Walker, L.P. (2006) Effect of cellulase mole fraction and cellulose recalcitrance on synergism in cellulose hydrolysis and binding. *Biotechnology Progress*, **22**, 270–277.
55. Nidetzky, B., Steiner, W., Hayn, M. & Claeysens, M. (1994) Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: A new model for synergistic interaction. *Biochemical Journal*, **298**, 705–710.
56. Reese, E.T. (1976) History of the cellulase program at the U.S. Army Natick Development Center. *Biotechnology and Bioengineering Symposium*, Wiley Interscience, New York, pp. 9–20.
57. Ghosh, A., Ghosh, B.K., Trimino-Vazquez, H., Eveleigh, D.E. & Montencourt, B.S. (1984) Cellulase secretion from a hypercellulolytic mutant of *Trichoderma reesei* RUT-C-30. *Archives of Microbiology*, **140**, 126–133.
58. Markov, A.V., Gusakov, A.V., Kondratyeva, E.G., Okunev, O.N., Bekkarevich, A.O. & Sinitsyn, A.P. (2005) New effective method for analysis of the component composition of enzyme complexes from *Trichoderma reesei*. *Biochemistry (Moscow)*, **70**, 657–663.
59. Yuan, S., Wu, Y. & Cosgrove, D.J. (2001) A fungal endoglucanase with plant cell wall extension activity. *Plant Physiology*, **127**, 324–333.
60. Rosgaard, L., Pedersen, S., Cherry, J.R., Harris, P. & Meyer, A.S. (2006) Efficiency of new fungal cellulase systems in boosting enzymatic degradation of barley straw lignocellulose. *Biotechnology Progress*, **22**, 493–498.
61. Saloheimo, A., Henrissat, B., Hoffren, A.M., Teلمان, O. & Penttilä, M. (1994) A novel, small endoglucanase gene, *egl5*, from *Trichoderma reesei* isolated by expression in yeast. *Molecular Microbiology*, **13**, 219–228.
62. Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyssönen, E., Bhatia, A., Ward, M. & Penttilä, M. (2002) Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials. *European Journal of Biochemistry*, **269**, 4202–4211.

63. Hu, J.P., Lanthier, P., White, T.C., McHugh, S.G., Yaguchi, M., Roy, R. & Thibault, P. (2001) Characterization of cellobiohydrolase I (Cel7A. glycoforms from extracts of *Trichoderma reesei* using capillary isoelectric focusing and electrospray mass spectrometry. *Journal of Chromatography. B, Biomedical Sciences and Applications*, **752**, 349–368.
64. Harrison, M.J., Nouwens, A.S., Jardine, D.R., Zachara, N.E., Gooley, A.A., Nevalainen, H. & Packer, N.H. (1998) Modified glycosylation of cellobiohydrolase I from a high cellulase-producing mutant strain of *Trichoderma reesei*. *European Journal of Biochemistry*, **256**, 119–127.
65. Kredics, L., Antal, Z., Szekeres, A., Hatvani, L., Manczinger, L., Vagvolgyi, C. & Nagy, E. (2005) Extracellular proteases of *Trichoderma* species. A review. *Acta Microbiologica et Immunologica Hungarica*, **52**, 169–184.
66. Hu, J.P., White, T.C. & Thibault, P. (2002) Identification of glycan structure and glycosylation sites in cellobiohydrolase II and endoglucanases I and II from *Trichoderma reesei*. *Glycobiology*, **12**, 837–849.
67. Aro, N., Saloheimo, A., Ilmen, M. & Penttilä, M. (2001) ACEII, a novel transcriptional activator involved in regulation of cellulase and xylanase genes of *Trichoderma reesei*. *Journal of Biological Chemistry*, **276**, 24309–24314.
68. Schulein, M. (1997) Enzymatic properties of cellulases from *Humicola insolens*. *Journal of Biotechnology*, **57**, 71–81.
69. Munoz, I.G., Ubhayasekera, W., Henriksson, H., Szabo, I., Pettersson, G., Johansson, G., Mowbray, S.L. & Stahlberg, J. (2001) Family 7 cellobiohydrolases from *Phanerochaete chrysosporium*: Crystal structure of the catalytic module of Cel7D (CBH58) at 1.32 Å resolution and homology models of the isozymes. *Journal of Molecular Biology*, **314**, 1097–1111.
70. Takada, G., Kawasaki, M., Kitawaki, M., Kawaguchi, T., Sumitani, J., Izumori, K. & Arai, M. (2002) Cloning and transcription analysis of the *Aspergillus aculeatus* No. F-50 endoglucanase 2 (cmc2) gene. *Journal of Bioscience and Bioengineering*, **94**, 482–485.
71. Grassick, A., Murray, P.G., Thompson, R., Collins, C.M., Byrnes, L., Birrane, G., Higgins, T.M. & Tuohy, M.G. (2004) Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus, *Talaromyces emersonii*. *European Journal of Biochemistry*, **271**, 4495–4506.
72. Bukhtojarov, F.E., Ustinov, B.B., Salanovich, T.N., Antonov, A.I., Gusakov, A.V., Okunev, O.N. & Sinitsyn, A.P. (2004) Cellulase complex of the fungus *Chrysosporium lucknowense*: Isolation and characterization of endoglucanases and cellobiohydrolases. *Biochemistry (Moscow)*, **69**, 542–551.
73. Ohtsuki, T., Yazaki, S., Ui, S., Mimura, A. & Suyanto (2005) Production of large multienzyme complex by aerobic thermophilic fungus *Chaetomium* sp. nov. MS-017 grown on palm oil mill fibre. *Letters in Applied Microbiology*, **40**, 111–116.
74. Tomme, P., Kwan, E., Gilkes, N.R., Kilburn, D.G. & Warren, R.A. (1996) Characterization of CenC, an enzyme from *Cellulomonas fimi* with both endo- and exoglucanase activities. *Journal of Bacteriology*, **178**, 4216–4223.
75. Lykidis, A., Mavromatis, K., Ivanova, N., Anderson, I., Land, M., Dibartolo, G., Martinez, M., Lapidus, A., Lucas, S., Copeland, A., Richardson, P., Wilson, D.B. & Kyrpides, N. (2007) Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. *Journal of Bacteriology*, **189**, 2477–2486.
76. Irwin, D.C., Cheng, M., Xiang, B., Rose, J.K.C. & Wilson, D.B. (2003) Cloning, expression and characterization of a family-74 xyloglucanase from *Thermobifida fusca*. *European Journal of Biochemistry*, **270**, 3083–3091.
77. Irwin, D.C., Jung, E.D. & Wilson, D.B. (1994) Characterization and sequence of a *Thermomonospora fusca* xylanase. *Applied and Environmental Microbiology*, **60**, 763–770.
78. McGrath, C.E. & Wilson, D.B. (2006) Characterization of a *Thermobifida fusca* beta-1,3-glucanase (Lam81A) with a potential role in plant biomass degradation. *Biochemistry*, **45**, 14094–14100.
79. Mosher, F., Irwin, D.C. & Wilson, D.B. (In press) *Biotechnology and Bioengineering*.

80. Spiridonov, N.A. & Wilson, D.B. (1998) Regulation of biosynthesis of individual cellulases in *Thermomonospora fusca*. *Journal of Bacteriology*, **180**, 3529–3532.
81. Lin, E. & Wilson, D.B. (1988) Transcription of the *celE* gene in *Thermomonospora fusca*. *Journal of Bacteriology*, **170**, 3838–3842.
82. Spiridonov, N.A. & Wilson, D.B. (1999) Characterization and cloning of CelR, a transcriptional regulator of cellulase genes from *Thermomonospora fusca*. *Journal of Biological Chemistry*, **274**, 13127–13132.
83. Spiridonov, N.A. & Wilson, D.B. (2001) Cloning and biochemical characterization of BglC, a β -glucosidase from cellulolytic actinomycete *Thermobifida fusca*. *Current Microbiology*, **42**, 295–301.
84. Taylor, L.E., II, Henrissat, B., Coutinho, P.M., Ekborg, N.A., Hutcheson, S.W. & Weiner, R.M. (2006) Complete cellulase system in the marine bacterium *Saccharophagus degradans* strain 2-40T. *Journal of Bacteriology*, **188**, 3849–3861.
85. Xie, G., Bruce, D.C., Challacombe, J.F., Chertkov, O., Giolna, P., Han, C.S., Henrissat, B., Lucas, S., Misra, M., Myers, G.L., Richardson, P., Tapia, R., Thayer, N.N., Thompson, L.S., Brettin, T.S., Wilson, D.B. & McBride, M.J. (2007) Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Applied Environmental Microbiology*, **73**, 3536–3545.
86. Iyo, A.H. & Forsberg, C.W. (1996) Endoglucanase G from *Fibrobacter succinogenes* S85 belongs to a class of enzymes characterized by a basic C-terminal domain. *Canadian Journal of Microbiology*, **42**, 934–943.
87. Malburg, S.R., Malburg, L.M., Jr., Liu, T., Iyo, A.H. & Forsberg, C.W. (1997) Catalytic properties of the cellulose-binding endoglucanase F from *Fibrobacter succinogenes* S85. *Applied and Environmental Microbiology*, **63**, 2449–2453.
88. Fields, M.W., Mallik, S. & Russell, J.B. (2000) *Fibrobacter succinogenes* S85 ferments ball-milled cellulose as fast as cellobiose until cellulose surface area is limiting. *Applied Microbiology and Biotechnology*, **54**, 570–574.
89. Cho, K.H. & Salyers, A.A. (2001) Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. *Journal of Bacteriology*, **183**, 7224–7230.
90. Zhang, S. & Wilson, D.B. (1997) Surface residue mutations which change the substrate specificity of *Thermomonospora fusca* endoglucanase E2. *Journal of Biotechnology*, **57**, 101–113.
91. Wolfgang, D.E. & Wilson, D.B. (1999) Mechanistic studies of active site mutants of *Thermomonospora fusca* endocellulase E2. *Biochemistry*, **38**, 9746–9751.
92. Zhang, S. & Wilson, D.B. (2000) Effects of non-catalytic residue mutations on substrate specificity and ligand binding of *Thermobifida fusca* endocellulase Cel6A. *European Journal of Biochemistry*, **267**, 244–252.
93. Barr, B.K., Wolfgang, D.E., Piens, K., Claeysens, M. & Wilson, D.B. (1998) Active-site binding of glycosides by *Thermomonospora fusca* endocellulase E2. *Biochemistry*, **37**, 9220–9229.
94. Larsson, A.M., Bergfors, T., Dultz, E., Irwin, D.C., Roos, A., Driguez, H., Wilson, D.B. & Jones, T.A. (2005) Crystal structure of *Thermobifida fusca* endoglucanase Cel6A in complex with substrate and inhibitor: The role of tyrosine Y73 in substrate ring distortion. *Biochemistry*, **44**, 12915–12922.
95. Andre, G., Kanchanawong, P., Palma, R., Cho, H., Deng, X., Irwin, D., Himmel, M.E., Wilson, D.B. & Brady, J.W. (2003) Computational and experimental studies of the catalytic mechanism of *Thermobifida fusca* cellulase Cel6A (E2). *Protein Engineering*, **16**, 125–134.
96. Nerinckx, W., Desmet, T. & Claeysens, M. (2003) A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of all glycoside hydrolases. *FEBS Letters*, **538**, 1–7.
97. Zhang, S., Irwin, D.C. & Wilson, D.B. (2000) Site-directed mutation of non-catalytic residues of *Thermobifida fusca* exocellulase Cel6B. *European Journal of Biochemistry*, **267**, 3101–3115.

98. Zhou, W., Irwin, D.C., Escovar-Kousen, J. & Wilson, D.B. (2004) Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes. *Biochemistry*, **43**, 9655–9663.
99. Li, Y., Irwin, D.C. & Wilson, D.B. (2007) Processivity, substrate binding and mechanism of cellulose hydrolysis of *Thermobifida fusca* Cel9A. *Applied Environmental Microbiology*, **73**, 3165–3172.
100. Mulakala, C. & Reilly, P.J. (2005) *Hypocrea jecorina* (*Trichoderma reesei*). Cel7A as a molecular machine: A docking study. *Proteins*, **60**, 598–605.
101. Sakon, J., Adney, W.S., Himmel, M.E., Thomas, S.R. & Karplus, P.A. (1996) Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry*, **35**, 10648–10660.
102. Varrot, A., Schulein, M., Fruchard, S., Driguez, H. & Davies, G.J. (2001) Atomic resolution structure of endoglucanase Cel5A in complex with methyl 4,4II,4III,4IV-tetrathio- α -cellopentoside highlights the alternative binding modes targeted by substrate mimics. *Acta Crystallographica. Section D, Biological Crystallography*, **57**, 1739–1742.
103. Varrot, A. & Davies, G.J. (2003) Direct experimental observation of the hydrogen-bonding network of a glycosidase along its reaction coordinate revealed by atomic resolution analyses of endoglucanase Cel5A. *Acta Crystallographica. Section D, Biological Crystallography*, **59**, 447–452.
104. Kashima, Y., Mori, K., Fukada, H. & Ishikawa, K. (2005) Analysis of the function of a hyperthermophilic endoglucanase from *Pyrococcus horikoshii* that hydrolyzes crystalline cellulose. *Extremophiles*, **9**, 37–43.

Chapter 12

Cellulase Systems of Anaerobic Microorganisms from the Rumen and Large Intestine

Harry J. Flint

12.1 Introduction

Herbivorous mammals do not secrete digestive enzymes that are able to degrade the major structural polysaccharides of plant cell walls, but depend on the activities of symbiotic gut microorganisms to obtain energy from the plant material that makes up the bulk of their diet. Breakdown of plant cell wall material is mediated by anaerobic microbial communities that develop in the large intestine (caecum and colon) in the case of hind-gut fermentors such as horses and rabbits, while in ruminants this breakdown occurs largely in the foregut, in the reticulo-rumen. In both cases, the short chain fatty acid products of fermentation are absorbed and used as energy sources by the animal, but the foregut location of the rumen also allows the animal to take advantage of microbial protein, by digesting the microbial cells that pass into the acidic stomach.

The rumen is the site of highly efficient breakdown of the wide variety of plant material that forms the diet of grazing animals, and harbors a complex consortium of anaerobic microorganisms comprising bacteria, archaea, fungi, and protozoa (1). Plant polysaccharide-degrading enzymes are produced by a high proportion of these microorganisms, both eukaryotes and prokaryotes. Many rumen microorganisms may, however, be considered as secondary utilizers that exist by cross-feeding, and are largely dependent on other organisms for primary attack upon the more recalcitrant plant structures (2–5). The role of their enzyme systems and transport machinery is to scavenge soluble polysaccharides and oligosaccharides as they are released by other microorganisms from plant material. The primary degraders of plant cell wall material on the other hand are assumed to be those that are capable of tight attachment and that possess the enzymatic machinery to access this complex, insoluble substrate. The number of genuinely cellulolytic species identified among the rumen microbiota is relatively small, and their populations may often be underestimated because of the difficulty of recovering them in the substrate. Interest has centered on these primary plant cell-wall-degrading species because of their key roles in initiating substrate breakdown.

12.2 Cellulolytic and hemicellulolytic bacteria from the rumen

Three rumen bacterial species, in particular, were recognized from early cultural studies to be actively cellulolytic. Two *Ruminococcus* species, *R. flavefaciens* and *R. albus*, are representatives of the Gram-positive Clostridial cluster IV, while *Fibrobacter succinigenes* belongs to a divergent group of Gram-negative bacteria. Recent molecular work has confirmed the importance of all three species in the rumen ecosystem (6–8) although there are indications from 16S rRNA diversity studies that hitherto uncultured cellulolytic bacteria remain to be recovered from the rumen (9).

12.2.1 *Ruminococcus flavefaciens*

There is good evidence that the majority of plant cell-wall-degrading enzymes in *R. flavefaciens* are retained on the bacterial cell surface via a cellulosome-type multienzyme complex. The assembly of this complex depends on the specific interaction of dockerin and cohesin domains within the component proteins. The known structural components of this complex are encoded by the *sca* gene cluster, which was first described in *R. flavefaciens* 17 (10). A very similar gene cluster has now also been identified from the partial genome sequence of *R. flavefaciens* FD1 (11). In *R. flavefaciens* 17, the scaffolding protein ScaA carries three cohesins and a C-terminal dockerin that in turn binds to any one of seven cohesins present in the larger noncatalytic protein ScaB (12). Recent work has shown that the C-terminus of ScaB also contains an unusual type of dockerin domain that interacts with a single cohesin present in the small, cell-surface-anchored protein, ScaE (13). ScaE is covalently bound to the peptidoglycan of the cell surface at its C-terminus via a sortase-mediated mechanism (Figure 12.1). Meanwhile, a second small protein encoded by the *sca* cluster, ScaC, possesses a dockerin that binds to ScaA, but carries a single, distinctive cohesin. This has been described as an adaptor, because it recognizes a different type of dockerin from that recognized by ScaA cohesins (14). Dockerin-carrying enzymes from *R. flavefaciens* 17 have so far been shown to bind mainly to ScaA cohesins, although some divergent enzyme dockerins such as those from CesaA (Ce3B) and XynE (Xyn11E) have distinct binding specificities for which the corresponding cohesin partners remain unknown (12, 15, 16).

The proteins encoded by the homologous *sca* cluster of *R. flavefaciens* FD1 differ in a few interesting respects from their homologues in *R. flavefaciens* 17. ScaA_{FD1} carries only two cohesins, while ScaB_{FD1} carries nine, five homologous to those of ScaB₁₇, and four closer to those of ScaA₁₇. Thus, ScaB_{FD1} is predicted to be able to bind four enzyme subunits directly, and 10 more via attached ScaA molecules, while ScaB₁₇ is predicted to bind 21 enzyme subunits via ScaA molecules (11). It is not yet clear whether these differences have functional consequences. Significant strain diversity exists in *R. flavefaciens* both at the genetic level and at the level of plant cell wall degradation (17, 18).

Most plant cell wall hydrolases studied from *R. flavefaciens* 17 have a dockerin-like module located either at the C-terminus, or internally (15, 19). One non-dockerin containing extracellular hydrolase is, however, known that carries family 11 and 10 xylanases domains at its N and C terminus, respectively, connected by an unusual NQ-rich linker (20). In addition to at least three types of enzyme dockerin, the dockerins found in the structural proteins ScaA

Cellulosome organization in *Ruminococcus flavefaciens* 17

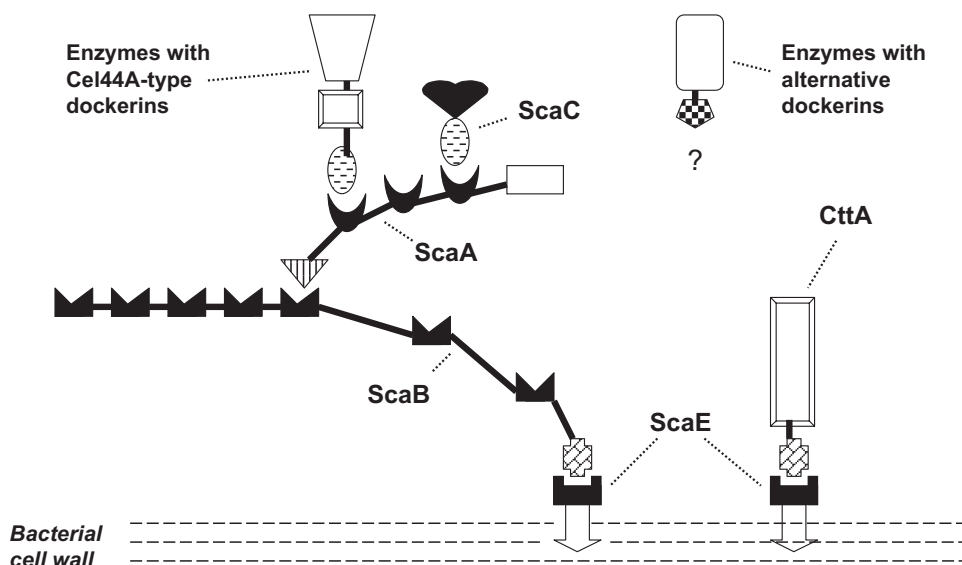


Figure 12.1 Postulated cellulosome structure in *Ruminococcus flavefaciens* 17. Additional cohesin-carrying proteins, yet to be identified, may be involved in interactions with enzymes possessing divergent dockerins.

and ScaB are also distinct in their sequences and binding characteristics, suggesting that there are at least five different dockerin specificities in this strain (21). Meanwhile, genome sequencing of *R. flavefaciens* FD1 has revealed at least 180 polypeptides that carry dockerin sequences (22). The sequences of these dockerins fall into a number of branches in phylogenetic analyses. Fewer than 50% of these carry glycoside hydrolase domains, and many carry proteinase domains or domains of unknown function. It remains unclear whether all these domains are required primarily for plant cell wall breakdown. Proteinases may well be involved in removing structural proteins in the plant cell wall, but it is also possible that dockerin-cohesin interactions mediate the assembly of protein complexes that play other roles on the bacterial cell surface.

Uniquely among known cellulosome systems, the *R. flavefaciens* scaffolding proteins so far identified carry no identified carbohydrate-binding modules. Carbohydrate-binding modules occur in many enzyme subunits, including a module in the abundant cellulase EndB (Cel44A) that represents a new CBM family. The gene that precedes ScaE, however, is now known to encode a protein (CttA) that is attached to the bacterial cell surface via ScaE, and binds crystalline cellulose (23). CttA is, therefore, a strong candidate for a carbohydrate-binding protein that may play a key role in positioning cellulosomal enzymes close to their substrates. The possible role played in adhesion by the yellow affinity substance of *R. flavefaciens* (24) remains unclear.

12.2.1.1 *Ruminococcus albus*

Recent analysis of two major cellulases from the proteome of *R. albus* 8 showed that both carried cellulose-binding modules, but neither possessed an obvious dockerin sequence (25). These enzymes, therefore, appear to be non-cellulosomal, although they might still be retained on the cell surface by other mechanisms. Meanwhile, a number of enzymes from *R. albus* 8 were also found to carry partially homologous C-terminal sequences that show remarkably broad polysaccharide-binding specificities (26). Although these sequences have potential roles in binding to dietary substrates, it is not ruled out that they might mediate binding of the polypeptides to carbohydrates on the cell surface, thus providing an anchoring mechanism. Other enzymes from *R. albus* 8 and from several other *R. albus* strains have, however, been reported to possess dockerins (27) and a large family of dockerin-containing proteins has been detected in the genome of *R. albus* 8 (28). The role and importance of cellulosome organization in this species, therefore, seems less clear than in *R. flavefaciens*, and this attribute may even vary between strains of *R. albus*.

Based on studies in two *R. albus* strains (8 and 20) type IV pili have been proposed to be involved in binding cells to cellulose (29–31). The pilA1 protein is thought to be a major component of the pili of *R. albus* 20, while pilA2 is more likely to play a role in pilus synthesis (31).

12.2.2 *Other Clostridium-related anaerobic bacteria*

Other species of Gram-positive cellulolytic rumen bacteria are known to include *Eubacterium cellulosolvens* (32, 33). Some strains of *Butyrivibrio fibrisolvens* have also been reported to be cellulolytic, although most strains of this species are actively hemicellulolytic, but not cellulolytic (34–36). Interestingly, there is evidence for cell-associated enzyme complexes in *B. fibrisolvens* (37). Sortase-mediated anchoring of individual enzymes to the bacterial cell wall has also been reported for amylases from a human colonic *B. fibrisolvens* strain and from the related *Roseburia inulinivorans* (38). It is likely, however, that many other species whose primary niche is not in plant cell wall breakdown possess enzymes for the transport and utilization of oligosaccharides released from plant cell walls. *Selenomonas ruminantium*, for example, has a xylan utilization operon (39) and is involved in interactions with cellulolytic species (40). Even *Streptococcus bovis*, assumed to be primarily a starch-degrading species, was found to possess a mixed link β -glucanase (41). The role of this enzyme was postulated to be in gaining access to starch through the removal of β -glucan-rich walls of cereal endosperm cells.

12.2.2.1 *Fibrobacter succinogenes*

Fibrobacter succinogenes and *F. intestinalis* are specialized cellulolytic bacteria that belong to a divergent phylum of Gram-negative anaerobes (42). Significant *Fibrobacter* populations can be detected by molecular probing (8), although they are poorly represented in amplified 16S clone libraries (43). Genome sequencing of *F. succinogenes* S85 has identified 113 genes that have likely roles in plant cell wall degradation, including 40 cellulases and 29 xylanases (44). As in Gram-positive cellulolytic anaerobes such as the ruminococci, complex

organization involving multiple catalytic domain and substrate-binding modules is a feature of *F. succinogenes* enzymes (1, 45). Dockerin-like domains have not been detected, however, and the organization of cellulolytic enzymes on the cell surface remains unclear. Some enzymes apparently share a basic C-terminal region, but it has not been established whether this is involved in cell surface attachment. Gene complement and cellulolytic activity are apparently well conserved among strains related to *F. succinogenes* S85, but with evidence of significant sequence divergence in *F. intestinalis* and some *F. succinogenes*-related strains (44, 46, 47).

F. succinogenes apparently lack xylose isomerase activity (48) and fail to utilize the breakdown products of xylan, despite possessing an array of xylanases. *R. flavefaciens* strains vary in their ability to utilize xylo-oligosaccharides, and appear to vary in respect of their xylose isomerase genes (49). For these bacteria, xylanase activity, therefore, appears to be required primarily for degrading the matrix polysaccharides thus facilitating access to the glucan components of the plant cell wall.

12.2.2.2 *Prevotella* species

Members of the CFB phylum, which includes *Prevotella* and *Bacteroides*, account for a very significant proportion (>30%) of total rumen bacteria (43, 50–51). Many of these species appear to be unique to the rumen (52). None of the available isolates are known to be cellulolytic, but a number of species, notably *P. bryantii* and *P. ruminicola*, possess carboxymethylcellulases, hemicellulases, and pectinases (53). This suggests a role in utilizing products of plant cell wall breakdown released by primary degraders, as suggested by coculture studies that demonstrate cross-feeding of oligosaccharides (2, 3, 54).

Gene clusters concerned with polysaccharide utilization have been studied in *P. bryantii* B₁4. One six-gene operon encodes two glucanases and a mannanase (55). The main role of at least one of the *P. bryantii* glucanases may be in degrading β (1,3-1,4) glucans (56, 57). Another operon encodes a family 43 xylosidase/exoxylanase and a family 10 endoglucanase (58) and shows homology with an operon from *Bacteroides ovatus* that is essential for xylan utilization (59). Expression of this gene cluster in *P. bryantii* is upregulated in response to substrate availability (60) with xylo-oligosaccharides providing the induction signal (61). Most of the xylanase activity found in *P. bryantii* cells is released only upon cell disruption, suggesting that it is located in the periplasm or membranes (62). The situation resembles that for starch-degrading enzymes in the human gut bacterium, *Bacteroides thetaiotaomicron*, where limited hydrolysis is thought to occur at the outer membrane, followed by extensive hydrolysis in the periplasm (63). An unusual family 10 xylanase encoded by an unlinked gene XynC (64) that has a preference for large xylo-oligosaccharides is a possible candidate for the limited extracellular xylanase activity in *P. bryantii* B₁4.

Genome sequences are available for some non-rumen representatives of the CFB division. The human colonic strain, *Bacteroides thetaiotaomicron* 5482, has a large genome (6.26 Mb) that exhibits considerable redundancy with respect to polysaccharide-degrading enzymes (65). This species is not a primary degrader of plant cell walls, and does not possess cellulases or xylanases belonging to the two major GH families 10 or 11, but possesses multiple xylosidase and glucosidase genes. Genome sequencing of rumen *Prevotella* species is likely to reveal greater complexity of xylanases and related enzymes than is known at present.

12.3 Plant cell wall breakdown by eukaryotic microorganisms

12.3.1 *Rumen fungi*

Rumen anaerobic fungi, which are related to chytrids (66), have been shown to produce highly active cellulase systems (67) and have been the source of recombinant enzymes with high specific activities that have attracted interest for a variety of biotechnological applications [e.g. (68, 69)]. As in cellulolytic bacteria, individual enzyme structures show multidomain organization (70). Furthermore, many polysaccharidases from anaerobic fungi exhibit a proposed 40 amino acid docking domain that may be present in one, two, or three copies (71). This domain is cysteine-rich and shows no sequence homology with bacterial dockerins, but has been reported in enzymes from a range of anaerobic fungi including *Orpinomyces*, *Piromyces*, and *Neocallimastix frontalis* (72). There is evidence in *Piromyces equi* for an interaction between this docking domain and a scaffolding protein of 97 kDa (73). Recent evidence in *Neocallimastix frontalis*, however, also suggests that removal of the docking domain influences the activity and temperature optimum of the adjacent catalytic domains (74). The nature of the docking domain interaction with the putative scaffolding protein remains to be established.

The close sequence relationships between catalytic domains from cellulolytic anaerobic fungi and bacteria suggest that horizontal gene transfer between these two groups has played an important role in the evolution of their cellulase systems (75).

12.3.2 *Rumen protozoa*

Rumen protozoa are reported to account for up to 50% of the microbial biomass in the rumen, and it has long been suspected that certain species play a significant role in the breakdown of plant material. There has been uncertainty, however, whether protozoa produce their own cellulases and xylanases, or whether the activities that can be measured for the protozoal fraction of rumen contents are due to ingested bacteria. Rumen protozoa cannot currently be maintained pure culture. To resolve this issue, cDNA libraries have been constructed from polyA⁺ RNA (presumed to be of predominantly eukaryotic origin) from a protozoal-enriched fraction of rumen contents. In order to focus on individual protozoal species, material for extraction was initially obtained from sheep that carry a complete bacterial flora, but that had been de-faunated and mono-associated with single strains of protozoa. Screening for activity allowed the recovery of multiple cDNA clones that express CMCase or xylanase activity from *Polyplastron multivesiculatum* (76–78). Consistent with their protozoal origin, these genes show extremely biased codon usage, and AT-rich flanking regions, and fail to cross-hybridize with DNA of bacterial origin. Representatives of family 11 and 10 xylanases from rumen protozoa are generally reported that have rather simple domain structures, although the family 10 domain in the enzyme XynB is adjacent to a family 22 substrate-binding module (78).

Polysaccharidases in rumen protozoa are assumed to be expressed mainly within the food vacuoles, rather than being extra-cellular as in rumen bacteria and fungi. What consequences this has for their organization remains unknown, but it perhaps makes it more likely that they exist as soluble enzymes rather than as a cell-bound complex. Expression of

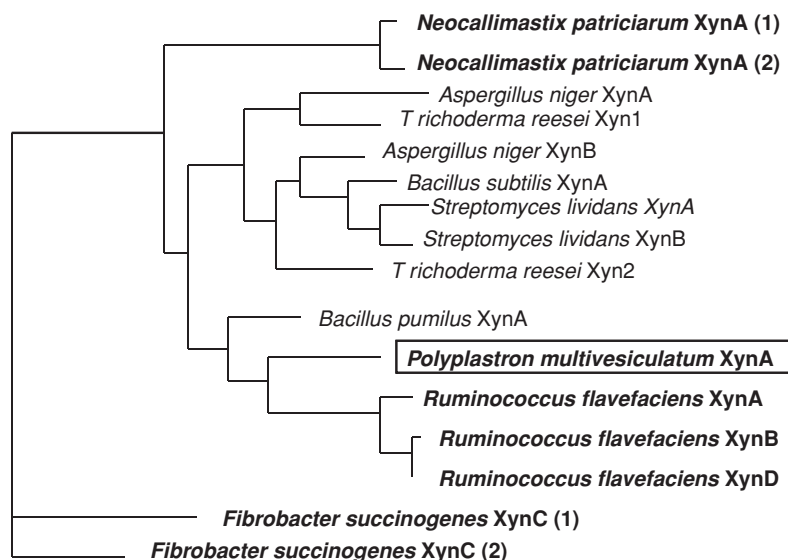


Figure 12.2 Phylogenetic relationship between a xylanase from the rumen protozoan *Polyplastron multivesiculatum* and other representative GH family 11 catalytic domains (based on Devillard *et al.* (77)). Sequences from rumen bacteria (*Fibrobacter*, *Ruminococcus*) and fungi (*Neocallimastix*) are shown in bold; those from non-rumen microorganisms are shown in normal type.

recombinant protozoal genes in *E. coli* may be reduced because of their highly biased codon usage.

The sequence relationships of protozoal glycoside hydrolases show that they are often quite closely related to bacterial enzymes (77, 79) (Figure 12.2). They are, therefore, strong candidates for acquisition by horizontal gene transfer. The dense, mixed communities of the gut create obvious potential for such transfer events, particularly in the case of protozoa that are constantly engulfing and digesting rumen bacteria.

12.4 Information from metagenomics

The extreme oxygen sensitivity and fastidious growth requirements of many rumen microorganisms mean that they may often be difficult to culture, or even unculturable. At the very least, the task of completing the description of rumen microbial diversity through cultivation looks daunting. An attractive alternative, therefore, is to build metagenomic libraries of rumen DNA that can be screened for activities of interest, or randomly sequenced to gain information on the genes present. This approach has been successfully applied to the recovery of genes encoding esterases, amylases and cellulases, xyloglucanases from the rumen (80, 81). Most of the glycoside hydrolases so far recovered appear to have simple domain structures, resembling those from the *Prevotella* group rather than those typical of known cellulolytic bacteria. This is most likely to reflect the relative ease of recovery of planctonic, rather than surface-attached bacteria, and perhaps also the relative ease of lysis

of Gram-negative cells. It may also reflect greater numbers of secondary utilizers of plant polysaccharides compared with primary degraders within the community. Nevertheless, the value of the approach is demonstrated by the fact that the collection includes representatives of new enzyme families. Information on sequence diversity can also be obtained without the need for library construction, either through amplification of specific genes by degenerate PCR, or directly by 454 sequencing.

12.5 The large intestine

Interest in lignocellulose breakdown in the gut has centered largely on the rumen, mainly because of the nutritional and economic importance of ruminal degradation, but also because of its relative accessibility as a gut microbial ecosystem. The large intestine of herbivores such as horses, however, harbor anaerobic communities of similar diversity and complexity (82) and include anaerobic bacteria, fungi, and protozoa. The major cellulolytic bacteria again appear to species of *Ruminococcus* and *Fibrobacter* (83). There is no indication that omnivores, including man, carry cellulolytic eukaryotes, but cellulolytic bacteria related to *Ruminococcus* spp. have been isolated (84). In addition, there is evidence that unknown *Clostridium*-related bacteria from human faecal samples can attach closely to wheat bran (85).

12.6 Conclusions

It is increasingly clear that the pool of DNA sequences coding for polysaccharide-degrading activities within the rumen microbial community has been subject to horizontal genetic exchange, including at some stage exchange between prokaryotic and eukaryotic microorganisms. The protein structures to which these catalytic and binding domains contribute, however, show an extraordinary diversity among rumen microorganisms. It seems safe to assume that the macromolecular organization of enzymes, adhesion mechanisms, and transport systems plays a crucial role in determining the ecological niche occupied by a given species. The details of this relationship, however, remain to be clarified.

Complex multidomain organization in polysaccharidases seems to be a feature of the primary cellulolytic species of anaerobic bacteria and fungi found in the rumen. At least in *R. flavefaciens* and cellulolytic rumen anaerobic fungi, a significant proportion of these polypeptides are organized into multienzyme cellulosome-type complexes that are anchored to the cell surface. While it is assumed that these complexes play a key role in cellulose and plant cell wall decomposition, as in cellulolytic *Clostridia* (86), functional evidence is still lacking. Furthermore, the mechanisms by which *R. albus* and *F. succinogenes*, two cellulolytic rumen bacteria that produce complex multidomain enzymes but have not been reported to display cellulosome organization, anchor their enzymes to the cell surface and achieve efficient plant cell wall breakdown remain unclear. Cellulolytic protozoa apparently achieve plant cell wall breakdown despite producing soluble enzymes of relatively simple structure, from the limited information currently available. These organisms may present a special case, however, since digestion is assumed to occur within food vacuoles that also contain ingested cellulolytic bacteria.

Analysis of completed genomes for representative cellulolytic microorganisms should soon provide crucial insights into their degradative enzyme systems. This information will need to be illuminated by functional studies, however, with a continuing requirement for gene transfer methodologies that can be applied to rumen microorganisms.

Acknowledgment

The Rowett Research Institute receives funding from the Scottish Executive Environment and Rural Affairs Department.

References

1. Chesson, A. & Forsberg, C.W. (1997) Polysaccharide degradation by rumen micro-organisms. In: *The Rumen Microbial Ecosystem* (eds P.N. Hobson & C.S. Stewart), pp. 329–381. Blackie Academic and Professional, London.
2. Dehority, B.A. (1991) Effects of microbial synergism on fiber digestion in the rumen. *Proceedings of the Nutrition Society*, **50**, 149–159.
3. Fondevila, M. & Dehority, B.A. (1996) Interactions between *Fibrobacter succinogenes*, *Prevotella ruminicola* and *Ruminococcus flavefaciens* in the digestion of cellulose from forages. *Journal of Animal Science*, **74**, 678–684.
4. Flint, H.J. (1997) The rumen microbial ecosystem – some recent developments. *Trends Microbiology*, **5**, 483–488.
5. Flint, H.J. (2004) Polysaccharide breakdown by anaerobic micro-organisms inhabiting the mammalian gut. *Advances in Applied Microbiology*, **56**, 89–120.
6. Krause, D.O., Bunch, R.J., Smith, W.J.M. & McSweeney, C.S. (1999) Diversity of *Ruminococcus* strains: A survey of genetic polymorphisms and plant digestibility. *Journal of Applied Microbiology*, **86**, 487–495.
7. Weimer, P.J., Waghorn, G.C., Odt, C.L. & Mertens, D.R. (1999) Effect of diet on populations of three species of ruminal cellulolytic bacteria in lactating dairy cows. *Journal of Dairy Science*, **82**, 122–134.
8. Michalet-Doreau, B., Fernandez, I. & Fonty, G. (2002) A comparison of enzymatic and molecular approaches to characterise the cellulolytic microbial ecosystems of the rumen and cecum. *Journal of Animal Science*, **80**, 790–796.
9. Larue, R., Yu, Z., Parisi, V.A., Egan, A.R. & Morrison, M. (2005) Novel microbial diversity adherent to plant biomass in the herbivore gastrointestinal tract, as revealed by ribosomal intergenic spacer analysis and rrs gene sequencing. *Environmental Microbiology*, **7**, 530–543.
10. Ding, S.Y., Rincon, M.T., Lamed, R., Martin, J.C., McCrae, S.I., Aurilia, V., Shoham, Y., Bayer, E.A. & Flint, H.J. (2001) Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *Journal of Bacteriology*, **183**, 1945–1953.
11. Jindou, S., Borovok, I., Rincon, M.T., Flint, H.J., Antonopoulos, D.A., Berg, M.E., White, B.A., Bayer, E.A. & Lamed, R. (2006) Conservation and divergence in cellulosome architecture between two strains of *Ruminococcus flavefaciens*. *Journal of Bacteriology*, **188**, 7971–7976.
12. Rincon, M.T., Ding, S.-Y., McCrae, S.I., Martin, J.C., Aurilia, V., Lamed, R., Shoham, Y., Bayer, E.A. & Flint, H.J. (2003) Novel organisation and divergent dockerin specificities in the cellulosome system of *Ruminococcus flavefaciens*. *Journal of Bacteriology*, **185**, 703–713.

13. Rincon, M.T., Cepelnik, T., Martin, J.C., Lamed, R., Bayer, E.A. & Flint, H.J. (2005) Unconventional mode of attachment of the *R. flavefaciens* cellulosome to the bacterial cell wall. *Journal of Bacteriology*, **187**, 7569–7578.
14. Rincón, M.T., Martin, J.C., Aurilia, V., McCrae, S.I., Rucklidge, G., Reid, M., Bayer, E.A., Lamed, R. & Flint, H.J. (2004) ScaC, an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the *Ruminococcus flavefaciens* 17 cellulosome. *Journal of Bacteriology*, **186**, 2576–2585.
15. Aurilia, V., Martin, J.C., McCrae, S.I., Scott, K.P., Rincon, M.T. & Flint, H.J. (2000) Three multidomain esterases belonging to the plant cell wall degrading enzyme system of the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17 carry divergent dockerin sequences. *Microbiology*, **146**, 1391–1397.
16. Rincon, M.T., McCrae, S.I., Kirby, J., Scott, K.P. & Flint, H.J. (2001) EndB, a multidomain cellulase from the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17, binds cellulose via a novel cellulose binding domain and to a 130 kDa *R. flavefaciens* protein via a dockerin domain. *Applied and Environmental Microbiology*, **67**, 4426–4431.
17. Krause, D.O., Dalrymple, B.P., Smith, W.J., Mackie, R.I. & McSweeney, C.S. (1999) 16S rDNA sequencing of *Ruminococcus albus* and *Ruminococcus flavefaciens*: Design of a signature probe and its application in adult sheep. *Microbiology*, **145**, 1797–1807.
18. Antonopoulos, D.A., Nelson, K.E., Morrison, M. & White, B.A. (2004) Strain-specific regions of *Ruminococcus flavefaciens* FD-1 as revealed by combinatorial random-phase genome sequencing and suppressive subtractive hybridization. *Environmental Microbiology*, **6**, 335–346.
19. Kirby, J., Martin, J.C., Daniel, A.S. & Flint, H.J. (1997) Dockerin-like sequences in cellulases and xylanases from the rumen cellulolytic bacterium *Ruminococcus flavefaciens*. *FEMS Microbiology Letters*, **149**, 213–219.
20. Zhang, J.-X. & Flint, H.J. (1992) A bifunctional xylanase encoded by the *xynA* gene of the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17 comprises two dissimilar domains linked by an asparagine/glutamine rich sequence. *Molecular Microbiology*, **6**, 1013–1023.
21. Flint, H.J. & Rincon, M.T. (2006) Cellulosome organisation in *Ruminococcus flavefaciens*. In: *Cellulosome* (eds. V. Umersky & I. Kataeva). Nova Science Publishers, New York.
22. Rincon, M.T., Borovok, I., Berg, M.E., Antonopoulos, D.S.A., Kim, R., Liu, L., Thimmapuram, J., Jindou, S., Lamed, R., Flint, H.J., Bayer, E.A. & White, B.A. (2006) Novel structural and catalytic elements in the *Ruminococcus flavefaciens* cellulosome revealed by genome analysis. *Reproduction, Nutrition and Development*, **46**(suppl 1), S57.
23. Rincon, M.T., Cepelnik, T., Martin, J.C., Barak, Y., Lamed, R., Bayer, E.A. & Flint, H.J. (submitted) A novel cell-surface anchored cellulose-binding protein encoded by the *sca* gene cluster of *Ruminococcus flavefaciens*. *Journal of Bacteriology*, **189**, 4774–4783.
24. Kopečný, J. & Hodrova, B. (1997) The effect of yellow affinity substance on cellulases of *Ruminococcus flavefaciens*. *Letters in Applied Microbiology*, **25**, 191–196.
25. Devillard, E., Goodheart, D.B., Sanjay, K.R., Karnati, S.K.R., Edward, A., Bayer, E.A., Lamed, R., Miron, J., Nelson, K.E. & Morrison, M. (2004) *Ruminococcus albus* 8 mutants defective in cellulose degradation are deficient in two processive endocellulases, Cel48A and Cel9B, both of which possess a novel modular architecture. *Journal of Bacteriology*, **186**, 136–145.
26. Xu, Q., Morrison, M., Nelson, K.E., Bayer, E.A., Atamna, N. & Lamed, R. (2004) A novel family of carbohydrate-binding modules identified with *Ruminococcus albus* proteins. *FEBS Letters*, **566**, 11–16.
27. Ohara, H., Karita, S., Kimura, T., Sakka, K. & Ohmiya, K. (2000) Characterisation of the cellulolytic complex (cellulosome) from *Ruminococcus albus*. *Bioscience Biotechnology Biochemistry*, **64**, 254–260.
28. Kang, S.H., Barak, Y., Lamed, R., Bayer, E.A. & Morrison, M. (2006) The functional repertoire of prokaryote cellulosomes includes the serpin superfamily of serine proteinase inhibitors. *Molecular Microbiology*, **60**, 1344–1354.

29. Pegden, R.S., Larson, M.A., Grant, R.J. & Morrison, M. (1998) Adherence of the Gram-positive bacterium *Ruminococcus albus* to cellulose and identification of a novel form of cellulose-binding protein which belongs to the Pil family of proteins. *Journal of Bacteriology*, **180**, 5921–5927.
30. Rakotoarivina, H., Jubelin, G., Hebraud, M., Gaillard-Martinie, B., Forano, E. & Mosoni, P. (2002) Adhesion to cellulose of the Gram-positive bacterium *Ruminococcus albus* involves type IV pili. *Microbiology*, **148**, 1871–1880.
31. Rakotoarivina, H., Larson, M.A., Morrison, M., Girardeau, J.P., Gaillard-Martinie, B., Forano, E. & Mosoni, P. (2005) The *Ruminococcus albus* pilA1-pilA2 locus: Expression and putative role of two adjacent pil genes in pilus formation and bacterial adhesion to cellulose. *Microbiology*, **151**, 1291–1299.
32. Stewart, C.S., Flint, H.J. & Bryant, M.P. (1997) The rumen bacteria. In: *The Rumen Microbial Ecosystem*, 2nd edn (eds P.N. Hobson & C.S. Stewart), pp. 10–72. Chapman and Hall, London.
33. Toyoda, A., Takano, K. & Minato, H. (2003) A possible role of cellulose-binding protein A (CbpA) in the adhesion of *Eubacterium cellulosolvens* 5 to cellulose. *Journal of General and Applied Microbiology*, **49**, 245–250.
34. Hespell, R.B. & Cotta, M.A. (1995) Degradation and utilization by *Butyrivibrio fibrisolvens* H17c of xylans with different chemical and physical properties. *Applied Environmental Microbiology*, **61**, 3042–3050.
35. McSweeney, C.S., Dulieu, A. & Bunch, R. (1998) *Butyrivibrio* and other xylanolytic organisms from the rumen have cinnamoyl esterase activity. *Anaerobe*, **4**, 57–65.
36. Dalrymple, B.P., Swadling, Y., Layton, I., Gobijs, K.S. & Xue, G.P. (1999) Distribution and evolution of the xylanase genes *xynA* and *xynB* and their homologues in strains of *Butyrivibrio fibrisolvens*. *Applied and Environmental Microbiology*, **65**, 3660–3667.
37. Lin, L.L. & Thomson, J.A. (1991) An analysis of the extracellular xylanases and cellulases of *Butyrivibrio fibrisolvens*. *FEMS Microbiology Letters*, **84**, 197–204.
38. Ramsay, A.G., Scott, K.P., Martin, J.C., Rincon, M.T. & Flint, H.J. (2006) Cell associated α -amylases of butyrate-producing Firmicute bacteria from the human colon. *Microbiology*, **152**, 3281–3290.
39. Whitehead, T.R. & Cotta, M.A. (2001) Identification of a broad specificity xylosidase/arabinofuranosidase important for xylo-oligosaccharide fermentation by the ruminal anaerobe *Selenomonas ruminantium*. *Current Microbiology*, **43**, 293–298.
40. Sawanon, S. & Kobayashi, Y. (2006) Synergistic fibrolysis in the rumen by cellulolytic *Ruminococcus flavefaciens* and *Selenomonas ruminantium*: Evidence in defined cultures. *Animal Science Journal*, **77**, 208–214.
41. Ekinci, M.S., McCrae, S.I. & Flint, H.J. (1997) Isolation and overexpression of a gene encoding an extracellular β (1,3-1,4) glucanase from *Streptococcus bovis* JB1. *Applied and Environmental Microbiology*, **63**, 3752–3756.
42. Griffiths, E. & Gupta, R.S. (2001) The use of signature sequences in different proteins to determine the relative branching order of bacterial division: Evidence that *Fibrobacter* diverged at a similar time to *Chlamydia* and the *Cytophaga-Flavobacterium-Bacteroides* division. *Microbiology*, **147**, 2611–2622.
43. Tajima, K., Aminov, R.I., Nagamine, T., Ogata, K., Nakamura, M., Matsui, H. & Benno, Y. (1999) Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. *FEMS Microbiology Ecology*, **29**, 159–169.
44. Qi, M., Nelson, K.E., Daugherty, S.C., Nelson, W.C., Hance, I.R., Morrison, M. & Forsberg, C.W. (2005) Novel molecular features of the fibrolytic intestinal bacterium *Fibrobacter intestinalis* not shared with *Fibrobacter succinogenes* as determined by suppressive subtractive hybridization. *Journal of Bacteriology*, **187**, 3739–3751.
45. Mitsumori, M., Xu, L., Kajikawa, H. & Kurihara, M. (2002) Properties of cellulose-binding modules in endoglucanase F from *Fibrobacter succinogenes* S85 by means of surface plasmon resonance. *FEMS Microbiology Letters*, **2154**, 277–281.

46. Bera-Maillet, C., Ribot, Y. & Forano, E. (2004) Fiber-degrading systems of different strains of the genus *Fibrobacter*. *Applied and Environmental Microbiology*, **70**, 2172–2179.
47. Flint, H.J., McPherson, C.A., Avgustin, G. & Stewart, C.S. (1990) Use of a cellulase encoding gene probe to reveal restriction fragment length polymorphisms among ruminal strains of *Bacteroides succinogenes*. *Current Microbiology*, **20**, 63–67.
48. Matte, A., Forsberg, C.W. & Gibbins, A. (1992) Enzymes associated with metabolism of xylose and other pentoses by *Prevotella* (Bacteroides) *ruminicola* strains, *Selenomonas ruminantium* D and *Fibrobacter succinogenes* S85. *Canadian Journal of Microbiology*, **38**, 370–376.
49. Aurilia, V., Martin, J.C., Munro, C.A., Mercer, D.K. & Flint, H.J. (2000) Organization and strain distribution of genes responsible for the utilization of xylans by *Ruminococcus flavefaciens*. *Anaerobe*, **6**, 333–340.
50. Whitford, M.F., Forster, R.J., Beard, C.E., Gong, J. & Teather, R.M. (1998) Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rDNA libraries. *Anaerobe*, **4**, 153–163.
51. Wood, J., Scott, K.P., Avgustin, G., Newbold, C.J. & Flint, H.J. (1998) Estimation of the relative abundance of different *Bacteroides* and *Prevotella* ribotypes in gut samples by restriction enzyme profiling of PCR-amplified 16S rRNA gene sequence. *Applied and Environmental Microbiology*, **64**, 3683–3689.
52. Ramsak, A., Peterka, M., Tajima, K., Martin, J.C., Wood, J., Johnston, M.E.A., Aminov, R.I., Flint, H.J. & Avgustin, G. (2000) Unravelling the genetic diversity of ruminal bacteria belonging to the CFB phylum. *FEMS Microbiology Ecology*, **33**, 69–79.
53. Avgustin, G., Wallace, R.J. & Flint, H.J. (1997) Phenotypic diversity among rumen isolates of *Prevotella ruminicola*: Proposal for redefinition of *Prevotella ruminicola* and the creation of *Prevotella brevis* sp. nov., *Prevotella bryantii* sp. nov. and *Prevotella albensis* sp. nov. *International Journal of Systematic Bacteriology*, **47**, 284–288.
54. Osborne, J.M. & Dehority, B.A. (1989) Synergism in degradation and utilization of intact forage cellulose, hemicellulose and pectin by three pure cultures of rumen bacteria. *Applied and Environmental Microbiology*, **55**, 2247–2250.
55. Gardner, R.G., Wells, J.E., Fields, M.W., Wilson, D.B. & Russell, J.B. (1997) A *Prevotella ruminicola* B₁4 operon encoding extracellular polysaccharide hydrolases. *Current Microbiology*, **35**, 274–277.
56. Fields, M.W., Russell, J.B. & Wilson, D.B. (1998) The role of ruminal carboxymethyl cellulases in the degradation of beta-glucans from cereal grain. *FEMS Microbiology Ecology*, **27**, 261–268.
57. Gasparic, A., Marinsek-Logar, R., Martin, J., Wallace, R.J., Nekrep, F.V. & Flint, H.J. (1995) Isolation of genes encoding xylanase, β -D-xylosidase and α -L-arabinofuranosidase activities from the rumen bacterium *Prevotella ruminicola* B₁4. *FEMS Microbiology Letters*, **125**, 135–142.
58. Gasparic, A., Daniel, A., Martin, J. & Flint, H.J. (1995) A xylan hydrolase gene cluster from *Prevotella ruminicola*: Sequence relationships, oxygen sensitivity and synergistic interactions of a novel exoxylanase. *Applied and Environmental Microbiology*, **61**, 2958–2964.
59. Weaver, J., Whitehead, T.R., Cotta, M.A., Valentine, P.C. & Salyers, A.A. (1992) Genetic analysis of a locus on the *Bacteroides ovatus* chromosome which contains xylan utilization genes. *Applied and Environmental Microbiology*, **58**, 2764–2770.
60. Miyazaki, K., Miyamoto, H., Mercer, D.K., Martin, J.C., Kojima, Y. & Flint, H.J. (2003) Involvement of the two component regulatory protein XynR in positive control of xylanase gene expression in the ruminal anaerobe *Prevotella bryantii* B₁4. *Journal of Bacteriology*, **185**, 2219–2226.
61. Miyazaki, K., Hirase, T., Kojima, Y. & Flint, H.J. (2005) Medium to large sized xylo-oligosaccharides are responsible for xylanase induction in *Prevotella bryantii* B₁4. *Microbiology*, **151**, 4121–4125.
62. Miyazaki, K., Martin, J.C., Marinsek-Logar, R. & Flint, H.J. (1997) Degradation and utilization of xylans by the rumen anaerobe *Prevotella bryantii* (formerly *P. ruminicola* subsp. *brevis* B₁4). *Anaerobe*, **3**, 373–381.

63. Anderson, K.L. & Salyers, A.A. (1989) Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer-membrane starch-binding sites and periplasmic starch-degrading enzymes. *Journal of Bacteriology*, **171**, 3192–3198.
64. Flint, H.J., Whitehead, T.R., Martin, J.C. & Gasparic, A. (1997) Interrupted domain structures in xylanases from two distantly related strains of *Prevotella ruminicola*. *Biochimica et Biophysica Acta*, **1337**, 161–165.
65. Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V. & Gordon, J.I. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science*, **299**(5615), 2074–2076.
66. Orpin, C.G. & Joblin, K.N. (1997) The rumen anaerobic fungi. In: *The Rumen Microbial Ecosystem* (eds P.N. Hobson & C.S. Stewart), pp. 140–195. Blackie, London, UK.
67. Wood, T.M., Wilson, C.A., McCrae, S.I. & Joblin, K.N. (1986) A highly active extracellular cellulase from the anaerobic rumen fungus *Neocallimastix frontalis*. *FEMS Microbiology Letters*, **34**, 37–40.
68. Xue, G.P., Johnson, J.S., Bransgrove, K.L., Gregg, K., Beard, C.E., Dalrymple, B.P., Gobius, K.S. & Aylward, J.H. (1997) Improvement of expression and secretion of a fungal xylanase in the rumen bacterium *Butyrivibrio fibrisolvens* OB156 by manipulation of promoter and signal sequences. *Journal of Biotechnology*, **54**, 139–148.
69. Ekinici, M.S., Martin, J. & Flint, H.J. (2002) Expression of a *celA* gene from the rumen fungus *Neocallimastix patriciarum* in *Streptococcus bovis* JB1 by means of promoter fusions. *Biotechnology Letters*, **24**, 735–741.
70. Dalrymple, B.P., Cybinski, D.H., Layton, I., McSweeney, C.S., Xue, G.P., Swadling, Y.J. & Lowry, J.B. (1997) Three *Neocallimastix patriciarum* esterases associated with the degradation of complex polysaccharides are members of a new family of hydrolases. *Microbiology*, **143**, 2605–2614.
71. Fanutti, C., Ponyi, T., Black, G.W., Hazlewood, G.P. & Gilbert, H.J. (1995) The conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from anaerobic fungi functions as a protein docking domain. *Journal of Biology and Chemistry*, **270**, 29314–29322.
72. Steenbakkers, P.J.M., Li, X., Ximenes, E.A., Arts, J.G., Chen, H.Z., Ljungdahl, L.G. & Op den Camp, H.J.M. (2001) Noncatalytic docking domains of cellulosomes of anaerobic fungi. *Journal of Bacteriology*, **183**, 5325–5333.
73. Fillingham, I.J., Kroon, P.A., Williamson, G., Gilbert, H.J. & Hazlewood, G.P. (1999) A modular cinnamoyl ester hydrolase from the anaerobic fungus *Piromyces equi* acts synergistically with xylanase and is part of a multienzyme cellulose-binding cellulose-hemicellulase complex. *Biochemistry Journal*, **343**, 215–224.
74. Huang, Y.-H., Huang, C.-T. & Hsue, R.-S. (2005) Effects of dockerin domains on *Neocallimastix frontalis* xylanases. *FEMS Microbiology Letters*, **243**, 455–460.
75. Gilbert, H.J., Hazlewood, G.P., Laurie, J.I., Orpin, C.G. & Xue, G.P. (1992) Homologous catalytic domains in a rumen fungal xylanase: Evidence for gene duplication and prokaryotic origin. *Molecular Microbiology*, **6**, 2065–2072.
76. Takenaka, A., D'Silva, C.G., Kudo, H., Itabashi, H. & Cheng, K.J. (1999) Molecular cloning, expression, and characterization of an endo-beta-1,4-glucanase cDNA from *Epidinium caudatum*. *Journal of General and Applied Microbiology*, **45**, 57–61.
77. Devillard, E., Newbold, C.J., Scott, K.P., Forano, E., Wallace, R.J., Jouany, J.P. & Flint, H.J. (1999) A family 11 xylanase from the ruminal protozoan *Polyplastron multivesiculatum*. *FEMS Microbiology Letters*, **181**, 145–152.
78. Devillard, E., Bera-Maillet, C., Flint, H.J., Scott, K.P., Newbold, C.J., Wallace, R.J., Jouany, J.P. & Forano, E. (2003) Characterisation of XynB, a modular xylanase from the ruminal protozoan *Polyplastron multivesiculatum*, bearing a family 22 carbohydrate binding module that binds to cellulose. *Biochemical Journal*, **373**, 495–503.
79. Ricard, G., McEwan, N.R., Dulith, B.A., Josalowski, T., Nhnotuany, J.P., Macheboeuf, D., Mitsumri, M., McIntosh, F.M., Michalowski, T., Nagamine, T., Nelson, N., Newbold, C.J., Nsabimana, E.,

- Takenaka, A., Thomas, N.A., Ushida, K., Hackstein, J.H.P. & Huynen, M.A. (2006) Horizontal gene transfer from bacteria to rumen ciliates indicates adaptation to their anaerobic, carbohydrate-rich environment. *BMC Genomics*, **7**, (22).
80. Ferrer, M., Golyshina, O.V., Chermikova, T.N., Kachane, A.N., Reyes-Duarte, D., Martins Dos Santos, V.A.P., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M.M., Timmis, K.N. & Golyshin, P.N. (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology*, **7**, 1996–2010.
81. Rincon, M.T., Henderson, D., Martin, J.C., Scott, K.P. & Flint, H.J. (2006) Polysaccharidases isolated from the rumen metagenome. *Reproduction, Nutrition and Development*, **46**(suppl 1), S57.
82. Daly, K., Stewart, C.S., Flint, H.J. & Shirazi-Beechey, S.P. (2001) Bacterial diversity within the equine large intestine as revealed by molecular analysis of cloned 16S rRNA genes. *FEMS Microbiology Ecology*, **38**, 141–151.
83. Julliand, V., de Vaux, A., Millet, L. & Fonty, G. (1999) Identification of *Ruminococcus flavefaciens* as the predominant cellulolytic bacterial species of the equine cecum. *Applied and Environmental Microbiology*, **65**, 3738–3741.
84. Robert, C. & Bernalier-Donadille, A. (2003) The cellulolytic microflora of the human colon: Evidence of microcrystalline cellulose-degrading bacteria in methane excreting subjects. *FEMS Microbiology Ecology*, **46**, 81–89.
85. McWilliam Leitch, E.C., Walker, A.W., Duncan, S.H., Holtrop, G. & Flint, H.J. (2006) Selective colonisation of insoluble substrates by human faecal bacteria. *Environmental Microbiology*, **9**, 667–679.
86. Bayer, E.A., Belaich, J.-P., Shoham, Y. & Lamed, R. (2004) The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology*, **58**, 521–554.

Chapter 13

The Cellulosome: A Natural Bacterial Strategy to Combat Biomass Recalcitrance

*Edward A. Bayer, Bernard Henrissat, and
Raphael Lamed*

13.1 Introduction

Plant cell wall polysaccharides provide an exceptional source of carbon and energy that can be potentially utilized as a low-cost renewable source of mixed sugars for fermentation to biofuels like ethanol (1–3). Perhaps the major bottleneck for conversion of biomass to ethanol is the combined high cost and low efficiency of the enzymes – the cellulases and other glycoside hydrolases – that are capable of degrading cellulose and myriads of complex plant cell wall polysaccharides to simple sugars. Efficient hydrolysis is impeded by limited accessibility of the enzymes and the recalcitrance of cellulose, owing to the extremely stable “paracrystalline” arrangement of the cellulose chains in the microfibrils (4).

The rate-limiting step in the hydrolysis of cellulose is not the catalytic cleavage of the β -1,4-glucosidic bond, but the disruption of a single chain of the substrate from its native crystalline matrix, thereby rendering it accessible to the active site of the enzyme. Thus, the processes and interactions that are most significant are those that facilitate the disruption of strong interchain hydrogen-bonding network that characterizes the microcrystalline arrangement of the insoluble cellulose structure (5). Moreover, single cellulolytic enzymes alone are generally incapable of efficient cellulose hydrolysis (6). The mode of action of the various cellulases is different, and they act synergistically, such that the combined extent of hydrolysis is much more than the sum of the individual parts (6–8). A simplistic example of this phenomenon is apparent when comparing the action of endoglucanases versus that of the exoglucanases: Endoglucanases cleave at random points along the cellulose chain, whereas exoglucanases cleave successive units (e.g., cellobiose or cellotetraose) from the chain ends in a “processive” (sequential) manner (9). Processive endoglucanases have also been described (10, 11), whereby an initial endo cleavage is followed by more systematic, successive cleavage along the chain, owing to either an appropriate change in the active-site architecture or the effects of an associated ancillary module(s) that might confer processivity on an otherwise endo-acting enzyme. In any case, the secret to potent enzymatic degradation of a recalcitrant cellulosic substrate is how these different types of enzymes work together, which is difficult to approach experimentally but is a key conceptual component for improving the overall degradation of cellulosic biomass.

In nature, degradation of cellulosic substrates is accomplished by various microorganisms, thus contributing a central component to the carbon cycle. In some cases, free-living microorganisms exploit such polysaccharides from decaying plant matter, found, for example, in compost piles and sewage sludge; in others, the microbes assist higher animals (e.g., ruminants, termites, etc.) in converting the polysaccharides to digestible components. Microbial degradation of cellulosic materials is one of the most important processes in nature, and different bacteria and fungi approach the task in different ways (12–16). Whereas aerobic microbes commonly produce copious quantities of the relevant enzymes (e.g., cellulases and hemicellulases), the biosynthetic apparatuses of anaerobes are much more frugal in their output of such enzymes. In this context, it is believed that the anaerobic environment presents a greater selective pressure for the evolution of highly efficient machinery for the extracellular degradation of polymeric substrates, such as the recalcitrant crystalline components of the plant cell wall. The energy yield of aerobes per hydrolyzed sugar unit is much greater than that of anaerobes, which have evolved extensive energy-conserving mechanisms for physiological adaptation to environmental stresses such as novel enzyme activities (17). Consequently, the anaerobes tend to adopt alternative strategies for degrading plant matter, and of these the organization of the enzymes into cellulosomes appears to be the most remarkable.

Much of the pioneering work on cellulases was first performed on “free” enzyme systems, commonly produced by fungi and some, aerobic bacterial species. The free cellulases are relatively simple macromolecules consisting generally of a catalytic module and cellulose-binding module (CBM) on a single polypeptide chain. The CBM targets the catalytic module to the cellulose surface, whereupon it begins to disrupt and degrade the cellulose chains. The different types of cellulase enzymes are thus distributed and interact freely in a synergistic manner (8).

13.2 The cellulosome concept

Fiber degradation is coordinated by a multitude of bacterial and fungal enzymes. In contrast to the free cellulase systems of the aerobic microorganisms, the particularly efficient cellulose degradation by anaerobic bacteria is achieved by the relatively small quantities of enzymes that they produce. This enigma was clarified in part by the identification of the cellulosome, produced by certain anaerobes: a multienzyme complex specialized in cellulose degradation, first studied in *Clostridium thermocellum* (13, 16, 18–26). The *C. thermocellum* cellulosomes reside primarily within cell surface protuberances and the culture medium (27–34).

The cellulosome comprises a set of multi-modular components – some structural and some enzymatic (see Figure 13.1). A pivotal noncatalytic subunit called *scaffoldin* secures the various enzymatic subunits into the complex, via the *cohesin–dockerin* interaction (19, 35–39). For this purpose, the primary scaffoldin possesses a series of functional modules, called cohesins, involved in enzyme attachment. In addition, the scaffoldin contains a cellulose-specific carbohydrate-binding module (CBM) for substrate targeting. The various enzyme subunits (notably, the cellulases and hemicellulases) each contain a specialized dockerin module, which is complementary to the scaffoldin-based cohesins. The specificity characteristics and tenacious binding between the scaffoldin-based *cohesin* modules and the enzyme-borne *dockerin* domains dictate the supramolecular architecture of the cellulosome.

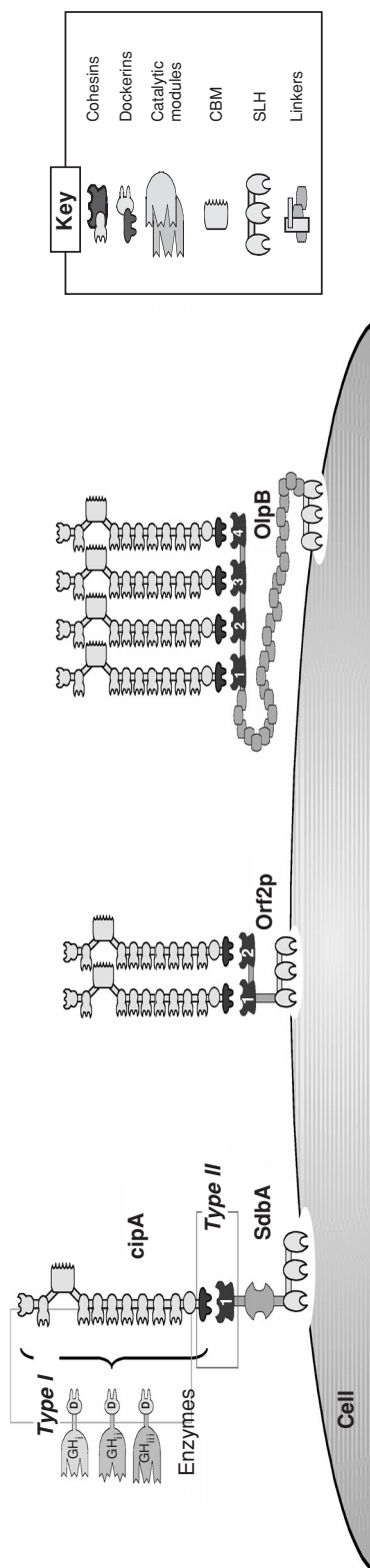


Figure 13.1 Current view of the *C. thermocellum* cellulosome and its attachment to the bacterial cell surface. The CBM borne by the CipA scaffoldin is believed to mediate the binding of the cell to the cellulose substrate and anchoring proteins SdbA, Orf2p, and OlpB mediate attachment of increasing numbers of CipA scaffoldins and their enzymes to the cell surface.

The organization of cellulases into a cellulosome is considered to be the most efficient of all microbial cellulose-degrading systems thus far studied (22, 23, 26). The cellulosomal scaffoldin of *C. thermocellum* also contains a dockerin of its own – a type-II dockerin that binds to type-II cohesins on one of three known anchoring scaffoldins (40–43). The latter contain one, two, or four type-II cohesins which incorporate the complementary number of scaffoldins onto the cell surface. The type-II dockerin is exquisitely specific for the type-II cohesins, and no interaction with its type-I counterpart takes place, thus ensuring correct cellulosome assembly and architecture (Figure 13.1).

The discovery and description of cellulosomes in different bacteria revealed marked diversity in the structure and architecture of their component parts. Although the existence of cellulosomes in a wide variety of different anaerobic bacteria has long been established (44), molecular details of various scaffoldin proteins, and, in particular, the interactions between the cohesin and dockerin, are only now beginning to emerge (35, 37, 45–52). There is also a limited amount of information that suggests that the composition and disposition of the cellulosome can be affected by carbon source (27, 53–62). Much remains to be learned about the extent to which cellulosome composition and location are affected by factors such as polysaccharide complexity and composition.

Recent work on *Ruminococcus flavefaciens* strain 17 has revealed an especially intricate cellulosome complex comprising numerous cohesin-containing scaffoldins, together with interacting enzymes and other unidentified dockerin-bearing proteins (Figure 13.2) (63–67). The assembly of these components differs markedly from the proposed molecular architecture in the *Clostridial* cellulosomes. In *R. flavefaciens*, ScaA incorporates a group of dockerin-containing enzymes into its three resident cohesin repeats (66, 67). In addition, a small ScaC scaffoldin serves as an adaptor protein that enhances the repertoire of cellulosomal subunits by binding both ScaA via its dockerin and to a range of as yet unidentified polypeptides via its single divergent cohesin (65). In turn, ScaA binds to any of seven cohesin repeats of ScaB via a specific cohesin–dockerin interaction, and ScaB is attached to the cell surface via a specialized cohesin–dockerin interaction with ScaE (68). ScaE includes an N-terminal cohesin and a C-terminal LPXTG-like motif, which suggests that it is positioned covalently on the cell surface via proteolytic cleavage and sortase-mediated attachment mechanism (69–72). This mechanism differs from the previously defined mode of cellulosome attachment in *C. thermocellum* via S-layer homology modules (41). Intriguingly, neither ScaA nor ScaB contains a CBM; it was thus unclear how the cellulosome or bacterial cell binds to cellulosic substrates. More recently, another important component of the *R. flavefaciens* sca gene cluster, *cttA*, has been shown to encode for a protein that also bears a C-terminal XDoc, which interacts with the ScaE cohesin (Rincon *et al.*, unpublished results). The gene product, CttA, contains two putative CBMs that bind to crystalline cellulose, thus providing, at least in part, insight into the molecular mechanism that accounts for the observed binding of the *R. flavefaciens* cell to cellulose. The system in *R. flavefaciens* therefore appears more intricate than those in *Clostridium* species (20, 22).

13.3 Cellulosomal carbohydrate-active enzymes

Initially, following their discovery, cellulosomes were thought to contain essentially enzymes involved in the degradation of cellulose, and the cellulosome was in fact originally

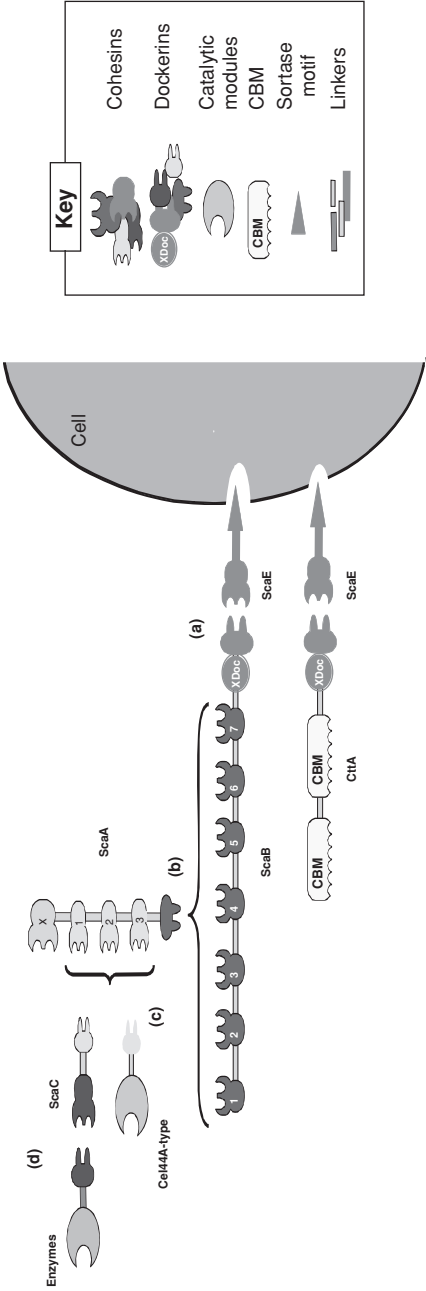


Figure 13.2 Schematic representation of our current view of cellulosome organization on the cell surface of *R. flavefaciens* strain 17. Four different cohesin-dockerin specificities have been identified (a–d).

defined as “a discrete cell surface organelle that exhibits cellulose-binding and various cellulolytic activities.” In the ensuing years, it became progressively clear that cellulosomes also contained non-cellulolytic enzymes (73–81). The first organism with cellulosomal components whose genome was completely sequenced was *Clostridium acetobutylicum* (82). Unfortunately, this genome did not reveal much on the cellulosome complexity as the *C. acetobutylicum* cellulosome is most likely ineffective or crippled (83, 84). The recent sequencing of three true cellulosome-producing bacteria *C. thermocellum* (http://genome.jgi-psf.org/draft_microbes/cloth/cloth.home.html), *R. flavefaciens* (85) and *C. cellulolyticum* (http://genome.jgi-psf.org/draft_microbes/cloce/cloce.home.html) allows us to grasp the sheer complexity of cellulosomes: these genomes encode a numbers of dockerin-containing proteins (from ~60 to ~170 depending on the bacterium), much larger than what a single scaffoldin can accommodate, with a *majority* of the enzymes being non-cellulolytic and specialized in the degradation of pectin and hemicelluloses. Each genome also reveals the presence of dockerin modules attached to proteins of completely unknown function.

Polysaccharide-cleaving enzymes (glycoside hydrolases and polysaccharide lyases) and accessory carbohydrate esterases are classified in a large number of families based on amino acid sequence similarities (86, 87), accessible online via the carbohydrate-active enzyme database (<http://www.cazy.org>). Our preliminary examination of the occurrence of members of these families in cellulosomes shows that in each of the three completely sequenced genomes mentioned above there is only one cellulosomal GH48 cellobiohydrolase, in sharp contrast to the occurrence of a much larger number of GH9 endoglucanases. In *C. thermocellum*, the GH48 enzyme is believed to be a major and decisive component of the cellulosome, when grown on microcrystalline cellulose (53, 88–93); growth of the bacterium on cellobiose results in reduced amounts of this component in the cellulosome (27, 53). The GH9 enzymes include several different modular themes and consequent alterations in activity patterns. The versatility in the repertoire of the cellulosomal GH9 enzymes may be a necessary and advantageous adaptation of cellulosomes. The various types of GH9 enzymes may undergo extensive regulation in response to changing product profiles, during the course of degradation of the plant cell wall. Interestingly, no member of family GH6 (candidate cellobiohydrolases) has been found in cellulosomal organisms so far, although they are common components of some aerobic cellulolytic bacteria and fungi.

In *C. thermocellum*, the intricacy of cellulosome assembly is further augmented by the complexity of some of its enzymes. Some of the cellulosomal enzymes are unusually large and contain numerous modules, and several contain more than one catalytic module (20, 21, 93–95). The ~180-kDa *C. thermocellum* CelJ, for example, contains six modules, including a GH9 and GH44 (96, 97); and CelH (~100 kDa) bears a GH5 and a GH26. The ~140-kDa family-9 CbhA also contains six modules (98). XynY and XynZ both contain a GH10 together with a family-1 carbohydrate esterase (a feruloyl esterase) (75, 99–101). Such bifunctional cellulosomal enzymes appear to be highly specialized for concerted cleavage of critical bonds of the native substrate. For example, the combination of a GH10 and feruloyl esterase in the xylanases would appear to be particularly effective in severing the covalent linkage between xylan and lignin in the plant cell wall. Why would the bacterium “choose” the solution of placing complementary activities here on the same polypeptide when it could have just incorporated them in the cellulosome? The answer to this probably sheds some light on the “proximity” effect in the cellulosome: If complementary enzymatic actions must occur on the *same* polymer chain, then the optimal solution is to place the enzymes

on the same polypeptide. The other obvious alternative would have been to just anchor the complementary activities on the same cellulosome; if the enzyme activities have to be in the same *microscopic* space, however, then the resultant cellulosome would probably be superior for degrading that particular microenvironment of the substrate (i.e., the plant cell wall). In addition, cellulosome composition can be further adjusted and refined by the bacterium, whereas the ratio between the catalytic activities in a bifunctional polypeptide cannot. The cellulosome can thus assemble, in a much more versatile manner, a larger number of catalytic subunits than that achieved by a single polypeptide.

Polysaccharide-degrading enzymes, whether hydrolases, lyases or accessory esterases, frequently contain appended carbohydrate-binding modules whose function is to target the enzyme to plant cell wall components (39, 102, 103). It has been suggested that bacteria have evolved different CBMs to target different polysaccharides or even different regions of a given polysaccharide (104). Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose (105). The number and the family membership of CBM-containing proteins in *C. thermocellum*, *R. flavefaciens*, and *C. cellulolyticum* is highly variable, but it is interesting to note that while some CBMs are common to cellulosomes and free bacterial systems (CBM3 and CBM6, for instance), cellulosomes so far appear completely devoid of CBM2 and CBM10 members, which are particularly abundant in the “free” cellulase-producing bacteria *Saccharophagus degradans* and *Cellvibrio japonicus* (Weiner, R.M., Taylor, L.E., II, Henrissat, B., Hauser, L., Land, M., Coutinho, P.M., Rancurel, R., Saunders, E., Longmire, A.G., Zhang, H., Bayer, E.A., Gilbert, H.J., Larimer, F., Jouline, I., Lamed, R., Richardson, P., Borovok, I. & Hutcheson, S.W. Complete genome sequence of the complex carbohydrate-degrading marine bacterium *Saccharophagus degradans* 2-40T. Submitted for publication; DeBoy, R.T., Tailford, L.E., Nagy, T., Topakas, V., Fouts, D.E., Mongodin, E.F., Emerson, J.B., Khouri, H., Henrissat, B., Gilbert, H.J. & Nelson, K.E. Insights into the mechanisms of plant cell-wall degradation from the genome sequence and transposon mutants of *Cellvibrio japonicus*. Submitted for publication). The absence of the family-2 and family-10 CBMs from the cellulosome repertoire remains unexplained. In addition, we currently have no explanation for the converse state: neither *S. degradans* nor *C. japonicus* appears to have any CBM3 module. The availability of the genome sequences of *C. thermocellum* and *C. cellulolyticum* on the JGI genome portal (http://genome.jgi-psf.org/mic_home.html) allows the first genome-wide comparison between these two authentic cellulosome-producing bacteria (Table 13.1). The preliminary analysis reveals an unexpected diversity in the CBMs of *C. cellulolyticum*, which, unlike *C. thermocellum*, has enzyme-bearing CBMs from families CBM9, CBM16, CBM17, and CBM28. On the other hand, *C. thermocellum* encodes a relatively large number of CBM3-containing proteins – twice that of *C. cellulolyticum* (Table 13.1). Although *C. cellulolyticum* encodes more glycoside hydrolases than *C. thermocellum*, the number of cellulosomal glycosidases (i.e., bearing dockerin modules) is very similar in both species. The small differences in the array of cellulosomal enzymes of the two species include the presence of candidate β -1,3-glucanases in *C. thermocellum* (families GH16 and GH81) and the presence of candidate α -galactosidase (GH27), β -galactosidase (GH59), α -L-arabinofuranosidase (GH62) and α -L-fucosidase (GH95) in *C. cellulolyticum*. These differences probably reflect differences in the particular ecology of each bacterium.

Outside of Clostridia there seem to be no trace of cellulosomes (assuming that they are traceable by the presence of dockerin modules; see below) or of their components.

Very few major confirmed oddities persist: (i) the archaeon *Archaeoglobus fulgidus* whose genome encodes two tandem ORFs with domains significantly similar to cohesins (106) and one dockerin-containing protein, and (ii) the bacterium *Gloeobacter violaceus* PCC 7421, whose genome encodes at least four dockerin-containing proteins (Genbank accession: BAC90876.1; BAC88000.1; BAC87999.1; BAC88001.1), but no cohesin (Coutinho & Henrissat, unpublished).

Is the presence of a dockerin-like sequence in a bacterial genome an indication of a cellulosome? Not necessarily. We have to remember that dockerins bear resemblance with calcium-binding EF-hands. In consequence, distant similarity to dockerin alone is no indication of a cellulosome structure. The case for a cellulosome would require the dockerin to be found attached to plant cell-wall-degrading enzymes along with the presence of a cohesin-containing protein. We tend to imagine that cellulosomes should all be like those of *C. thermocellum* because *C. thermocellum* was the first one found, the best studied, and closely related systems have been found (in closely related organisms). Of course one can imagine other aggregated systems whose aggregation would be promoted by other ways than the cohesin–dockerin interaction. For instance, the cellulolytic bacterium *Cytophaga hutchinsonii* does not have genes encoding proteins with cohesin domains, which are characteristic of cellulosomes. However, many proteins predicted to be involved in polysaccharide utilization have repeat domains which are thought to be involved in anchoring proteins to the cell surface (107).

13.4 The cellulosome–cellulose interaction

The binding of the cellulosome to microcrystalline cellulose has been suggested in the past to be mediated directly via the scaffoldin-based, cellulose-binding CBM (37, 38). In the original description of the cellulosome from *C. thermocellum*, a mutant was isolated that lacked cell-surface cellulosomes and consequently failed to bind to the substrate (27, 108). Since the scaffoldin is anchored to the cell surface via the various anchoring scaffoldins, the same CBM appears to be responsible for binding the entire bacterial cell to cellulosic substrates, as will be described below (109). Nevertheless, the *C. thermocellum* cellulosome also includes numerous enzymes that also bear family-3 CBMs (93, 95) that can potentially supplement the cellulose-binding role of the scaffoldin-based CBM.

The recombinant, scaffoldin-derived, cellulose-binding family-3 CBM was also shown to bind chitin but not xylan (110). The CBM binds to celluloses of very high crystallinity – even to cellulose of the highest known crystallinity, namely, from the cell walls of the algae, *Valonia ventricosa*. It also binds strongly to amorphous forms of cellulose, including phosphoric acid swollen cellulose of minimal crystallinity. The very high capacity to bind to the amorphous cellulose (111) likely reflects increased accessibility to the binding sites, rather than a preference for amorphous regions. In fact, the comparative crystal structures of different families of cellulose-binding CBMs suggest that, in each case, an aromatic strip of amino acids on the flat surface of the CBM molecule mediates strong binding with the glucose chain of the hydrophobic face of the cellulose surface (112). Additional hydrogen-bonding interactions between hydrophilic amino acids and polar groups on neighboring glucose chains are thought to provide additional protein–cellulose contacts that stabilize the strong interaction. Docking analyses thus suggest that a patch along three adjacent cellulose

chains interacts with the flat surface of the CBM molecule, which indicates why cellobiose and other cellodextrins fail to inhibit CBM binding to the insoluble substrate (110).

The *C. thermocellum* genome includes several other, enzyme-borne family-3 CBMs that exhibit most or all of the latter proposed binding residues (113–115), all of which would presumably bind strongly to microcrystalline cellulose. However, the latter five or six enzymes, lack dockerins and are thus free, non-cellulosomal enzymes. In addition, another sub-class of family-3 CBMs, termed CBM3c, is fused to a GH9 catalytic module. This type of family-3 CBM is modified in the same surface residues that are considered to bind crystalline cellulose, and the CBM3c is believed to assist the adjacent catalytic module in binding to structural intermediates of the substrate or to alter the mode of activity (113, 116, 117). The CBM3c fails to bind strongly to crystalline cellulose substrates, and thus plays an ancillary role in breakdown of the cellulose chain by the catalytic module. The balance of the other dockerin-containing, enzyme-borne CBMs lack several but not all of the cellulose-binding residues. In some cases (118, 119), cellulose-binding properties have been reported, but their contribution to the primary strong recognition of crystalline cellulose by the cellulosome, as observed for the scaffoldin-based family-3 CBM, remains unclear.

The binding of the cellulosome to cellulose is inhibited by low ionic strength, and water can be used to release at least part of a cellulosome or CBM preparation from the cellulose matrix (111). Increasing the salt content of the medium increases the binding of the cellulosome to the substrate. Interestingly, maximal enzymatic activity of the cellulosome complex was observed at suboptimal conditions of adherence to the substrate. At low salt concentrations, the low enzymatic activity most likely reflects the lack of sufficient adsorption, whereas above-optimum concentrations of salt, the reduced activities may be due to restricted mobility of the cellulosome under such conditions. Once attached, the cellulosome seems to remain static, as shown by laser bleaching experiments (unpublished results). A micrograph of cellulosomes bound to the surface of the super-crystalline *Valonia* cellulose is shown in Figure 13.3. Note the apparent cellulosome-induced stripping of the microfibrils from the surface of the cellulose crystal.

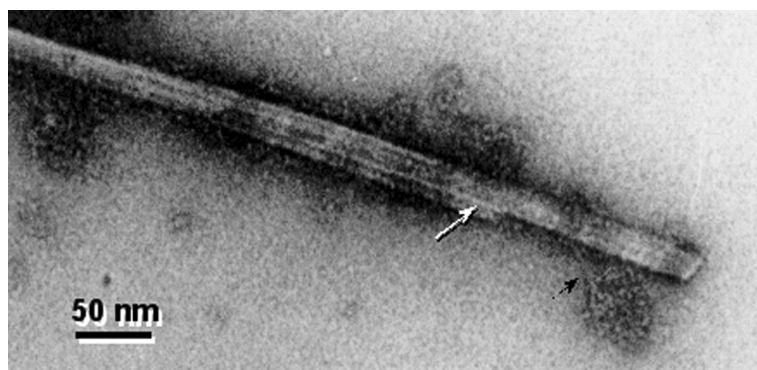


Figure 13.3 Transmission electron micrograph of purified *C. thermocellum* cellulosomes adsorbed onto cellulose microcrystals from the algae, *Valonia ventricosa*. Note the etched fibrils on the surface of the cellulose substrate (white arrow). A single microfibril seems to have been detached from the cellulose surface (black arrow) and is apparently being processed by the enzymes of an attached cellulosome complex. [Micrographs courtesy of Claire Boisset and Henri Chanzy (CNRS – CERMAV, Grenoble, France.)]

The organization of enzymes into cellulosomes ensures their multiple, concerted attachment to the substrate surface. This arrangement affords a distinct advantage over the distribution of free cellulases, since it allows multiple hydrolysis and facilitates access of additional appropriate cellulases and other enzymes (e.g., hemicellulases, pectate lyases, carbohydrate esterases) at or near the cleavage sites for enhanced processing of the substrate. The proximity of complementary enzymes at the cellulose surface provides a suitable remedy for counteracting substrate recalcitrance, as will be further discussed below.

13.5 Cell-surface disposition of cellulosomes

Early on, it was recognized that cellulosomes were intimately associated with the bacterial cell surface of *C. thermocellum* (88, 108, 120). As mentioned above, the molecular mechanism for this phenomenon was later determined. The type-II cohesin–dockerin complex fixes the primary scaffoldin and its complement of enzymes to one of the anchoring proteins, which also contains an S-layer homology (SLH) module that mediates attachment to the peptidoglycan of the cell surface (Figure 13.1) (36, 40–43, 121). Other species of bacteria, e.g., *Acetivibrio cellulolyticus* and *Bacteroides cellulosolvens*, possess anchoring protein(s) that exhibit similar type-II cohesins and SLH modules for attachment of the enzyme-bearing primary scaffoldin to the cell surface (88, 108, 120, 122–124). However, this is not the only strategy. As described above, ScaE of *R. flavefaciens* contains a sortase signal motif through which the scaffoldin is covalently attached to the peptidoglycan (Figure 13.2) (68). Other bacteria, notably *C. cellulolyticum* and *C. cellulovorans*, are also known to bear cellulosomes on their cell surface (125, 126), but there is as yet no evidence for involvement of an SLH module or sortase-mediated covalent attachment. The known scaffoldin of both of the latter species includes two and four copies of a hydrophilic domain (37, 46), which, in the case of *C. cellulovorans*, has been implicated in a cell-attachment function (127), but the true function of this particular domain awaits further experimental verification.

The attachment of the cellulosome to the cell surface, coupled with the CBM of the primary scaffoldin, inferred that the cellulosome is intimately involved in cell adhesion to the insoluble substrate. Indeed, the initial demonstration of the cellulosome in *C. thermocellum* was assisted by the isolation of an adherence defective mutant that was also impaired in its arrangement of cellulosomes on the cell surface (108). Cellulosomes are packaged into polycellulosome-containing protuberances (Figure 13.4), which undergo a dramatic conformational change upon interaction with the cellulose surface (27–29). The protuberances thus protract to form “contact corridors” laden with fibrous material that connect between cellulose-bound cellulosomes and the cell surface thus promoting cell uptake of the cellulose degradation products. The cellulosomal enzymes are subject to potent product inhibition (89, 111), and cell uptake and assimilation of the cellulose-degradation products (i.e., cellobiose and higher order soluble cellodextrins) relieves the inhibitory effect, thereby allowing facile hydrolysis of the insoluble substrate to proceed unhindered. Moreover, additional saccharolytic bacteria, which share the same ecosystem with the polymer-degrading strains (e.g., *C. thermocellum*), further purge the immediate environment of the inhibitory sugars, thus propagating substrate degradation even more (19, 128).

The cellulosome-producing cellulolytic bacterium *A. cellulolyticus* is characterized by an especially elaborate surface morphology; when grown on cellulosic substrates, the

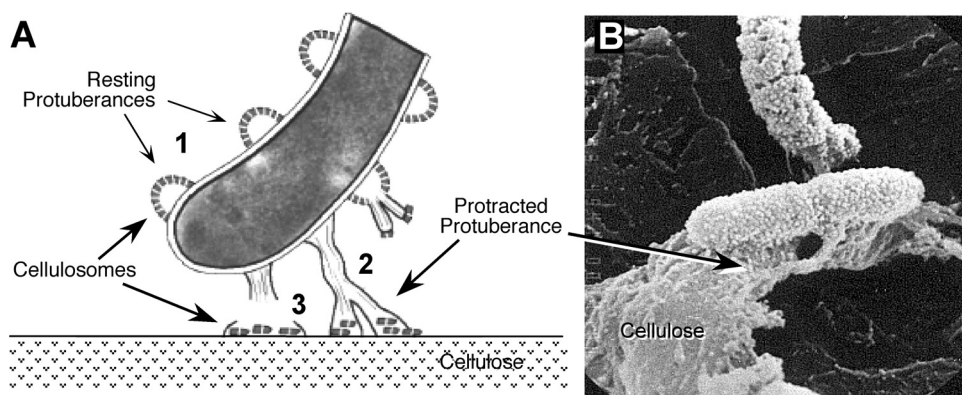


Figure 13.4 Proposed interaction of cellulosome-producing bacteria with cellulosic substrates. (Panel A) Schematic representation of the interaction of cell-surface cellulosome-laden protuberance-like organelles with the cellulose surface. 1) Prior to contact with the cellulosic substrate, the protuberances are compacted in a resting state; 2) upon cell-substrate contact, the protuberances protract whereby the cellulosomes interact directly with the cellulose surface; 3) eventually, the cell detaches from the substrate, leaving cellulosome clusters on the cellulose surface. (Panel B) Scanning electron microscopy (SEM) of *Acetivibrio cellulolyticus* bound to cellulose, showing the presence of large characteristic protuberance-like structures on the cell surface. The cellulose-bound cells appear to be connected to the substrate via structural extensions of the cell-surface protuberances.

cellulosome-bearing protuberances inundate the surface of the substrate, a process which appears to assist the cells in overcoming the inherent recalcitrance of the cellulose (Figure 13.4B). During stationary phase growth on cellulose, the cells detach from the substrate, leaving a legacy of attached cellulosomes that continue their hydrolytic action in the absence of the parent cell (18, 28, 129, 130).

13.6 Cellulosome assault on recalcitrant cellulose substrates

In studying the action of cellulases on cellulose, different model cellulosic substrates are employed, which exhibit different levels of crystallinity and accessibility to enzyme, two of the major parameters that contribute to the overall recalcitrance of the substrate. It should also be remembered that the procedures used for the preparation of these model cellulose substrates result in a form, which is very different from that of the native cellulose microfibrils within the plant cell wall.

The least crystalline substrates used for such studies are the soluble derivatized forms, such as carboxymethyl cellulose or hydroxymethyl cellulose, which, due to the substituted groups, are essentially non-crystalline in nature. The soluble, derivatized celluloses are used as a substrate for endoglucanases that can cleave along the cellulose chain, providing that the derivatization is not too extensive. Exoglucanases begin at the chain ends, and, upon meeting the substituent group, immediately come to a halt. Thus, such soluble derivatized substrates can be used to differentiate between the endo- and exo-acting enzymes. It should be noted that cellulases tend to exhibit a spectrum between the endo- and exo-modes of

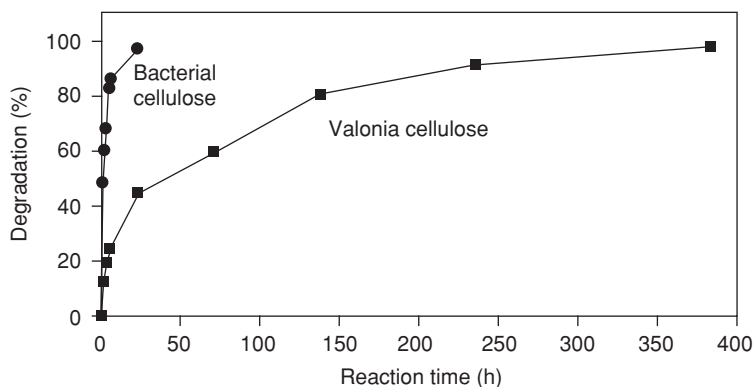


Figure 13.5 Comparative degradation bacterial cellulose ribbons versus *Valonia* cellulose microcrystals by the *C. thermocellum* cellulosome.

cellulolytic activity; and several enzymes are known to display an endo-processive action – i.e., the initial steps of cleavage may be endo-acting whereby the enzyme binds in some manner to the internal portion of the cellulose chain, whereupon the enzyme continues to degrade the substrate processively.

The next level of crystallinity is exemplified by phosphoric acid-treated celluloses, sometimes referred to as amorphous cellulose, sometimes acid-swollen cellulose. The crystallinity of the latter substrate depends on the conditions used for treatment and is usually very low, although residual levels of crystallinity may be observed. In any case, some enzymes, whether endo- or exo-acting, degrade this type of substrate readily, whereas others do not. In this case, the exact mechanistic features of substrate-versus-enzyme that generate extensive degradation are not understood.

Cotton cellulose is about 45% crystalline and is sometimes used as a model plant-derived substrate. Avicel, a commercial preparation of microcrystalline cellulose, is commonly used as a model substrate for examination of cellulolytic enzyme systems. Avicel is also about 45% crystalline (131, 132), but exhibits heightened recalcitrance owing to its large particle size (due to the drying of the material during its preparation) and consequent inaccessibility of the inner cellulose chains. Bacterial cellulose ribbons, produced by the bacterium *Acetobacter xylinum*, are about 65% crystalline, and BMCC (bacterial microcrystalline cellulose) is prepared by acid hydrolysis of amorphous regions of the bacterial cellulose ribbons, thus resulting in a crystallinity of about 70%. By far, however, the most crystalline and recalcitrant form of cellulose is derived from the *Valonia ventricosa* cell wall, which is close to 100% crystalline (131).

Despite the rather high crystallinity of bacterial cellulose ribbons, the *C. thermocellum* cellulosome completely dissolves this substrate relatively rapidly, within a 24-hour period under the conditions of the assay (Figure 13.5). Under the same conditions, a sluggish but relentless degradation of the highly recalcitrant *Valonia* cellulose is achieved, reaching near completion only after a 16-day time interval (133).

The modified morphologies of the different substrates can be followed ultrastructurally by transmission electron microscopy, and the images provide insight into the mechanism and

extent of degradation of recalcitrant substrates (134–137). Cellulosome-induced degradation of bacterial cellulose ribbons was indicative of a digestion pattern suggesting a concerted assault of different types of cellulases (e.g., combined endo- and exo-acting) on the cellulose substrate (Figure 13.6), consistent with the spatial proximity of the two types of enzymes in the cellulosomes (133). In this context, the observed cleavage of the cellulose ribbons is the signature of an endo mode of action, and the observed defibrillation of the substrate (see Figure 13.6B) corresponds to processive action associated with exo-acting cellulases. The residual bacterial cellulose, following 85% degradation (Figure 13.6C), bears no resemblance to the fine ribbons that were originally subjected to cellulosome action.

The images of partially degraded, highly recalcitrant *Valonia* cellulose show a rather different picture (Figure 13.7). Unlike the digestion pattern observed for the bacterial cellulose ribbons, the degradation of *Valonia* microcrystals is accompanied by distinctive digestive features, including crystal thinning and pointed tips, reminiscent of previously characterized features during the degradation of this substrate by individual and combined fungal cellulases. The thinning feature was previously characterized as a function of processive action by the exo-acting fungal cellobiohydrolase I (Cel7A), whereas pointed tips were associated with the action of the less processive “unidirectional” digestion (i.e., vis-à-vis non-reducing to reducing or vice versa) of cellobiohydrolase II (Cel6A) (135, 138).

Interestingly, the micrograph in Figure 13.7B shows both types of features. Even though over 95% of the cellulose have been digested, the individual residual cellulose microcrystals often show both features and some are apparently unchanged from the original images. It seems as if the individual cellulosomes from the same batch exhibit a wide diversity in their mode of action, probably related to the inherent heterogeneity in their enzyme content. The persistence of pointed tips, indicates unidirectional processivity (135, 138). Many of the properties of the intact cellulosome seem to be analogous to those of this particular processive enzyme. The direction of cleavage of the family-48 enzymes is from the reducing to non-reducing ends (139). The persistence of intact *Valonia* cellulose crystals, even after near-complete digestion may indicate that the rate-determining step in its degradation is the initial attack, once consummated, the crystal undergoes rapid degradation. Indeed, the rate-limiting step of cellulose degradation has been considered to be the separation of the individual cellulose chains from the crystal lattice. Once exposed, the battery of enzymes can deal both with the separated chain as well as the void left in the crystal.

13.7 Degradation of cellulose by the *C. thermocellum* cellulosome

The efficiency of a crude cell-free cellulase system from *C. thermocellum* for the biodegradation of crystalline cellulose was first reported by Johnson *et al.* (140). These authors claimed that the activity of this cellulase system toward cotton was at least 50 times higher than that of the extracellular cellulase system from *T. reesei*. This level of disparity has since been tempered somewhat, and elevated (~4-fold) levels of cellulose degradation have been estimated in favor of the cell-surface cellulases from *C. thermocellum* over the free cellulase system of *T. reesei* (16, 141). It is in fact very hard to assess this difference, since the equilibration and estimation of equivalent crude or isolated preparations of cellulases from two

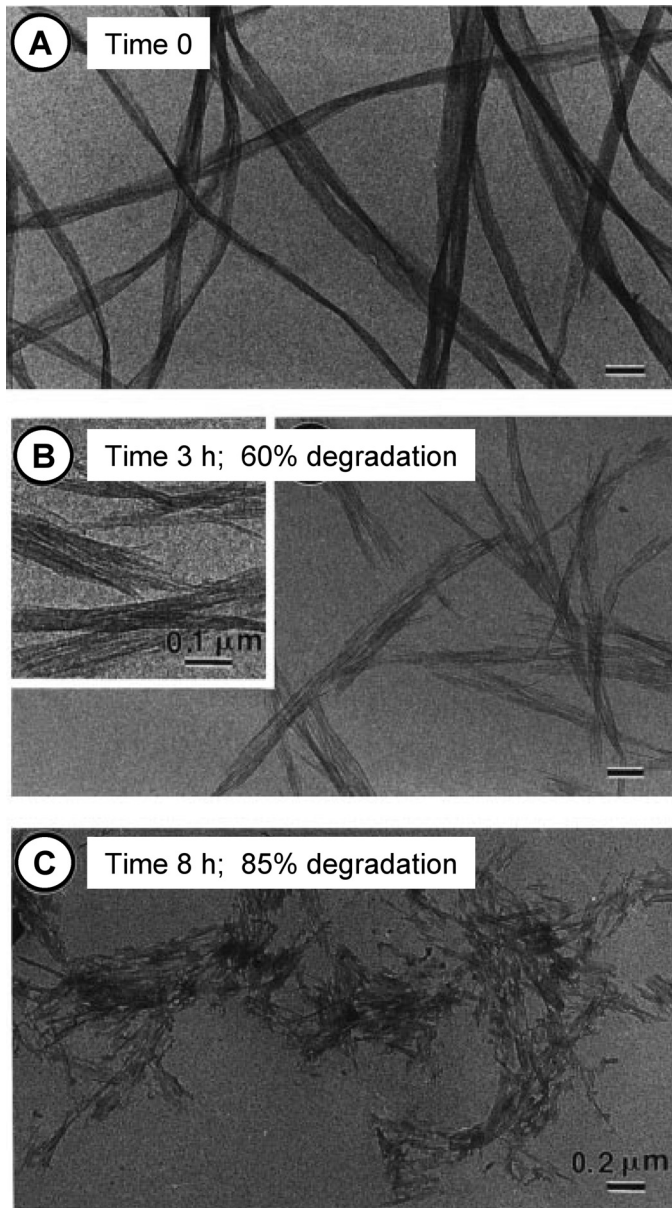


Figure 13.6 Transmission electron microscopy of cellulosome-induced degradation of bacterial cellulose ribbons. (A) Untreated substrate. (B) Bacterial cellulose following 3 hours of digestion using preparations of the *C. thermocellum* cellulosome. (C) As in (B), but after 6.5 hours of digestion.

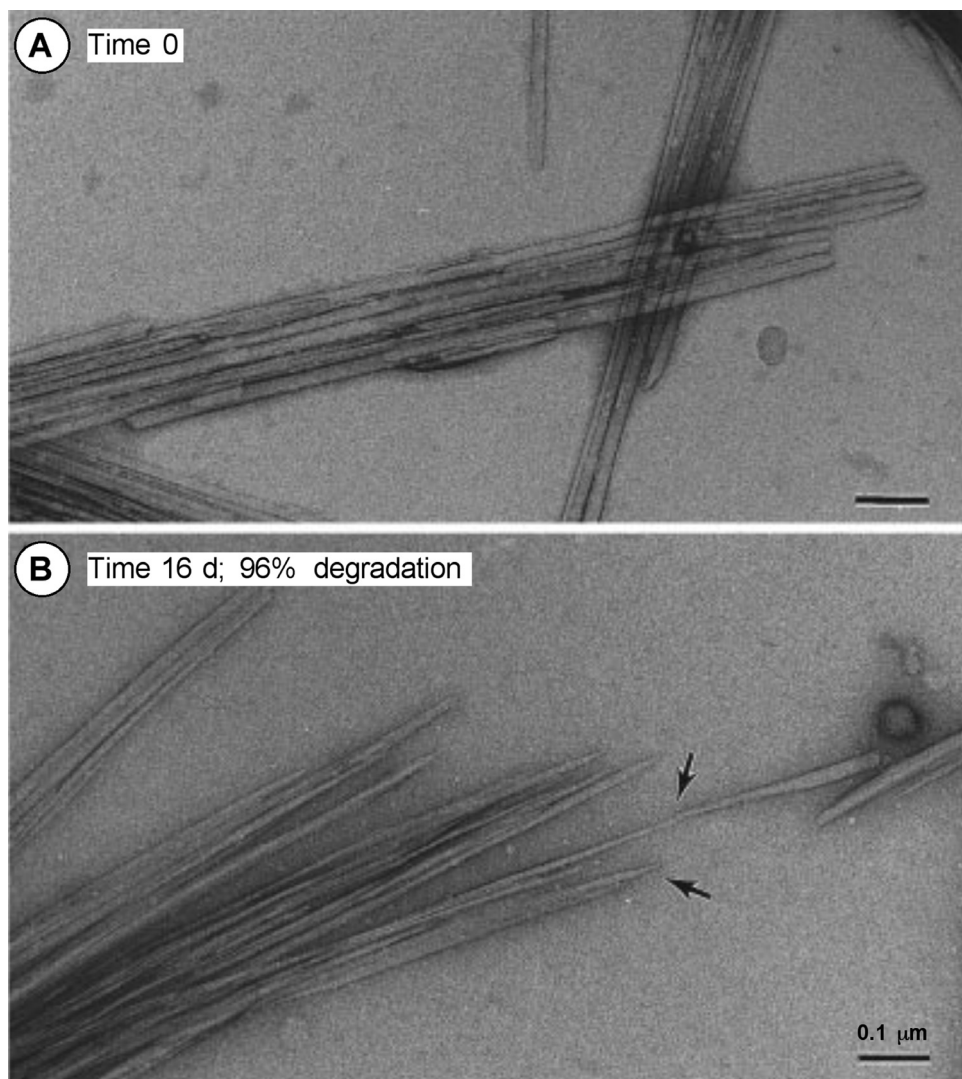


Figure 13.7 Transmission electron microscopy of cellulosome-induced degradation of *Valonia* cellulose microcrystals. (A) Untreated microcrystals. (B) *Valonia* microcrystals following 16 days of digestion using preparations of the *C. thermocellum* cellulosome. The arrows indicate pointed microcrystals, characteristic of the unidirectional action of exo-acting cellulase.

different species are difficult in themselves to determine. Nevertheless, such attempts have consistently shown that the *C. thermocellum* cellulosome is superior in its cellulolysis of recalcitrant cellulosic substrates when compared to the free fungal cellulase systems (Figure 13.8). The principal family-48 processive cellulase is decisive to the observed decomposition of the substrate, since cellulosome preparations that are deficient in this enzyme display reduced levels of hydrolysis on recalcitrant forms of cellulose (142). In any case, it is important

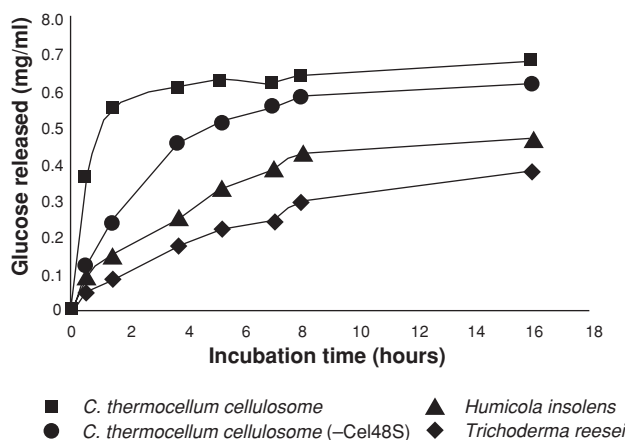


Figure 13.8 Cellulase activity of the *C. thermocellum* cellulosome versus those of the free cellulase systems from the fungi, *T. reesei* (*Hypocrea*) and *Humicola insolens*. With one exception, the cells were grown on microcrystalline cellulose (Avicel), and crude preparations of the extracellular enzymes (cellulosome or free cellulases) were produced. The curve (filled circles) denoting *C. thermocellum* (-Cel48S) represents a cellulosome preparation, derived from cells grown on cellobiose instead of cellulose, under conditions that result in highly reduced quantities of the Cel48S cellobiohydrolase in the cellulosome (27, 53, 142). Avicel was employed as a substrate and subjected to treatment using equivalent amounts of the bacterial cellulosome or fungal cellulase preparations.

to note that although bacterial cellulosomes seem to exhibit enhanced activity compared to that of the fungal enzymes, anaerobic bacteria produce much less cellulolytic enzymes (<1 g/L) than do the fungi (~100 g/L). In view of this imbalance, industry has consistently turned to the economically favorable fungal enzymes, which are thus preferred in all current industrial applications of cellulases.

The cellulolytic potential of the cellulosome on Avicel was in fact demonstrated many years ago (143). For these experiments, cellulosome action was enhanced by inclusion of the *Aspergillus niger* β -glucosidase, which served to counteract the inhibitory effect of cellobiose on the cellulosomal enzymes. The β -glucosidase severs the β -1,4 bond of cellobiose to produce two molecules of the non-inhibitory glucose product. Without this added enzyme, the course of cellulose hydrolysis by the cellulosome is rapidly impeded. In its presence, however, facile degradation of relatively low concentrations of cellulose in suspension (20 g/L) is achieved to completion in a relatively short time period (Figure 13.9). Complete digestion of concentrated cellulose suspensions (200 g/L) are also attained, provided that optimal amounts of cellulosome complex are included in the reactor (Figure 13.9, arrow).

13.8 The cellulosome rationale

Why did some cellulolytic anaerobes evolve to produce such a complicated mechanism in the production of an intricate multi-component conglomerate of enzymes, CBMs and other functional modules, complexed into a discrete type of complex, which is located in

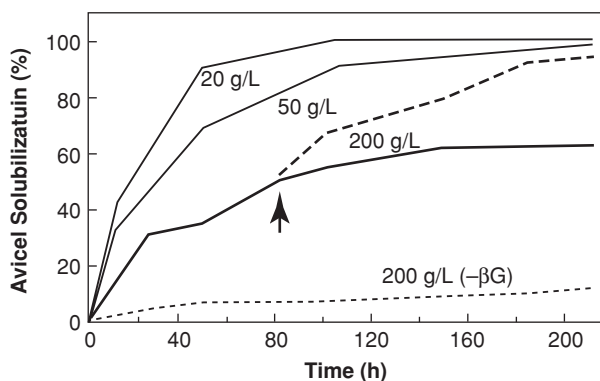


Figure 13.9 Solubilization of concentrated suspensions of microcrystalline cellulose by the combined cellulosome β -glucosidase system. The figures show the time course of solubilization of the indicated concentrations of Avicel using purified preparations of the *C. thermocellum* cellulosome (8 μ g per mg cellulose) together with the *A. niger* β -glucosidase (0.04 cellobiase units per mg substrate). At very high substrate concentrations, a second sample of the combined enzyme system was applied (arrow) to achieve near-complete solubilization of substrate (dashed line). Without the added β -glucosidase ($-\beta$ G), only meager levels of cellulose solubilization are observed.

bundle-like organelles over the cell surface? Such an expense in metabolic energy would imply that the bacterium would gain a significant compensation for the effort. A biochemical rationale for enhanced cellulolytic activity of cellulosomes on recalcitrant forms of cellulose was originally proposed already upon its discovery (88, 120):

The (cellulosome) complex apparently comprises various different forms of cellulases, each of which may bear separate specificities toward different quaternary structures on the complex cellulose substrate. The major organizational role of this complex might be designed for effective delivery to the substrate as well as to bring into proximity the various complementary enzymes (e.g., exo and endocellulases). In addition, the complex may be structured in such a way as to enable the protection of various product intermediates and to facilitate their transfer to other cellulase components for further hydrolysis. In any event, the cellulase subunits seem to be arranged within the CBF (cellulosome) complex in a defined supramolecular fashion designed for highly efficient cellulose degradation.

Proof of the targeting and proximity effects was eventually realized through the use of artificial “designer cellulosomes,” which enabled controlled incorporation of selected cellulases into a cellulosome-like complex (144, 145). For this purpose, a chimeric scaffoldin that contains divergent cohesins and matching dockerin-bearing enzymes can be mixed in vitro to form minicellulosomes of defined composition and spatial arrangement (Figure 13.10). The capacity to prepare cellulosomes of uniform composition and to control the types of enzymes included therein has contributed a better understanding of the factors important for efficient cellulosome action. Thus, the two major factors that serve to enhance deconstruction of recalcitrant forms of cellulose are indeed targeting to the substrate surface by the scaffoldin-borne cellulose-binding module (CBM), and the consequent proximity of the enzyme components.

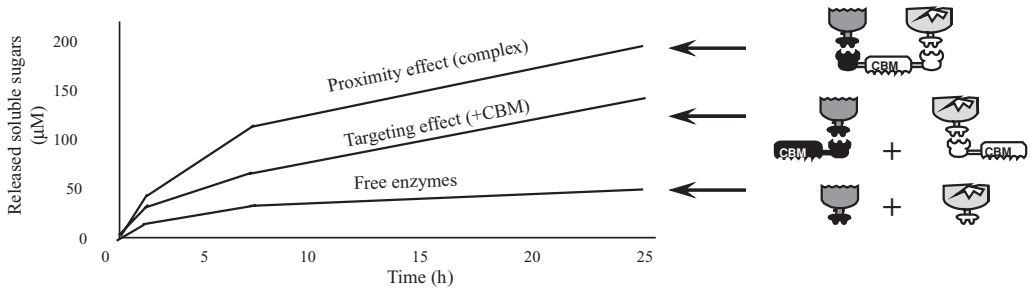


Figure 13.10 Enhanced synergy of binary designer cellulosomes, by combined targeting and enzyme proximity effects (145). Hybrid enzymes, each bearing a divergent dockerin but lacking a CBM, act synergistically to degrade microcrystalline cellulose. Employment of a matching cohesin-CBM construct serves to target the enzymes to the substrate, thereby resulting in enhanced synergistic action. Integration of the enzymes into a single complex via a chimeric binary scaffoldin generates an additional enhancement of synergy.

The incorporation of enzymes into cellulosomes would ensure that different complementary activities are all contained *in the same microscopic area* of the cellulose substrate rather than being dispersed statistically over its surface. Separation of relatively small amounts of enzymes by large distances would thwart the synergistic action of complementary enzymes, each of which bears a CBM that prevents its free diffusion as in the case of the free (non-cellulosomal) enzymes. The free systems are limited by the fact that such enzymes are statistically isolated and hence do not benefit of the full complementary action of the other enzymes. The “proximity effect” of the cellulosomes counteracts this problem. The “remedy” for the free systems is the production of very large quantities of enzymes to overcome their solitude, which is the strategy taken by aerobic bacterial and fungal systems. In fact, this may be the reason why cellulose-binding CBMs are appended to non-cellulolytic enzymes: both cellulosomes and the free enzyme systems use CBMs to prevent the useless dispersion of their enzymes (whether cellulolytic or not). Disruption of the sites of the cellulose-hemicellulose interface is a key to potent digestion of the plant cell wall. It is imperative that the different enzymes, the cellulases and hemicellulases, are in close proximity to each other. Because of its large size (relative to bacteria), the fungi probably deliver their cellulase mixture in a localized specific area (tip of hypha perhaps) and then the enzymes remain together in that area as a function of their respective CBMs. In contrast, due to its smaller size, it is *C. thermocellum*’s best interest to remain attached to a multienzyme system capable of accommodating almost any type of enzyme needed to breakdown completely a piece of plant cell wall *without having to move*. This may be the reason that cellulosomes appear to be so superior to the free enzyme systems.

With the exception of the cellulosome cluster in *C. acetobutylicum*, which probably produces a crippled complex (83, 84), it can be seen from the few known genomes of cellulosome-producing bacteria that there is a much larger number of dockerin-containing enzymes that can be attached to a scaffoldin than there are slots (cohesins) on the scaffoldin. In these bacteria, there are more hemicellulose-degrading enzymes than there are cellulases. This infers that cellulosome organization is beneficial *not only* for crystalline cellulose digestion, but

equally for the digestion of hemicelluloses, pectins, and other plant cell wall polysaccharides. This holds true even for *C. thermocellum*, a bacterium capable of assimilating only cellulose and its degradation products. Nevertheless, it appears to serve as the major polysaccharide-degrading bacterium in its ecosystem, passing on the surplus (e.g., cellobiose) and/or superfluous (e.g., xylose, etc.) sugars to its companion (saccharolytic) strains that share the same locale. Particular types of plant cell wall-derived biomass require very precise enzyme mixtures. In this context, previous attempts to “benchmark” one cellulase “system” against another did not particularly reflect an intrinsic ability but the “fit” of the enzyme mixture with respect to the actual substrate used. In other words, it is hard to compare one complex mixture to another.

For many decades, it was believed that the solution to efficient degradation of recalcitrant cellulose would be found in conventional engineering approaches by employing a set of soluble enzymes, such as the then-known endo- and exo-glucanases. The amount of financial resources versus the expectations of society resulted in great disappointment that served to minimize further scientific activity in this type of research. Very little has been achieved on the scientific front during the interim period. However, some positive results in this area have been recorded, including the discovery of the cellulosome (the topic of this chapter) and the realization that the bacterial cellulases are multi-modular entities that exhibit distinct functionalities. It is hoped that in the future these novel contributions will encourage a new burst of serious scientific and applied efforts in this area. The resolution and harnessing of the cellulosome (146) may thus provide a renewed opportunity for combating the difficulties encountered in digesting recalcitrant cellulosic biomass in an effective, cost-efficient manner.

Acknowledgments

The authors are grateful to Claire Boisset, Henri Chanzy, Pedro M. Coutinho, the late Martin Schüle, Ely Morag, Yuval Shoham, Ilya Borovok, Henri-Pierre Fierobe, Jean-Pierre and Anne Belaich, Marco Rincon and Harry Flint for their input, discussions, and collaboration, past and present. This work was funded by Research Grants 394/03 and 422/05 from the Israel Science Foundation (Jerusalem), a grant from the Alternative Energy Research Initiative, and by grants from the United States–Israel Bi-national Science Foundation (BSF), Jerusalem, Israel. EAB is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry at The Weizmann Institute of Science.

References

1. Schubert, C. (2006) Can biofuels finally take center stage? *Nature Biotechnology*, **24**, 777–784.
2. Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Jr., Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R. & Tschaplinski, T. (2006) The path forward for biofuels and biomaterials. *Science*, **311**, 484–489.
3. DOE U.S. (2006) *Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda*, DOE/SC-0095. U.S. Department of Energy Office of Science and Office of Energy Efficiency and Renewable Energy.

4. Himmel, M.E., Ding, S.-Y., Johnson, D.K., Adney, W.S., Nimlos, M.R., Brady, J.W. & Foust, T.D. (2007) Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*, **315**, 804–807; Erratum: 316, 982.
5. Matthews, J.F., Skopeck, C.E., Mason, P.E., Zuccato, P., Torget, R.W., Sugiyama, J., Himmel, M.E. & Brady, J.W. (2006) Computer simulation studies of microcrystalline cellulose I β . *Carbohydrate Research*, **341**, 138–152.
6. Reese, E.T., Siu, R.G.H. & Levinson, H.S. (1950) The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *Journal of Bacteriology*, **59**, 485–497.
7. Wood, T.M. & McCrae, S.I. (1978) Cellulase of *Trichoderma koningii*—purification and properties of some endoglucanase components with special reference to their action on cellulose when acting alone and in synergism with cellobiohydrolase. *Biochemical Journal*, **171**, 61–72.
8. Henrissat, B., Driguez, H., Viet, C. & Schülein, M. (1985) Synergism of cellulases from *Trichoderma reesei* in the degradation of cellulose. *Bio/Technology*, **3**, 722–726.
9. Irwin, D., Walker, L., Spezio, M. & Wilson, D. (1993) Activity studies of eight purified cellulases: Specificity, synergism, and binding domain effects. *Biotechnology and Bioengineering*, **42**, 1002–1013.
10. Reverbel-Leroy, C., Pagés, S., Belaich, A., Belaich, J.-P. & Tardif, C. (1997) The processive endocellulase CelF, a major component of the *Clostridium cellulolyticum* cellulosome: Purification and characterization of the recombinant form. *Journal of Bacteriology*, **179**, 46–52.
11. Wilson, D.B., Irwin, D., Sakon, J. & Karplus, P.A. (1998) *Thermomonospora fusca* cellulase E4: A processive endocellulase. In: *Carbohydrases from Trichoderma reesei and Other Microorganisms* (eds M. Claeysens, W. Nerinckx & K. Piens), pp. 133–138. The Royal Society of Chemistry, London.
12. Béguin, P. & Aubert, J.-P. (1994) The biological degradation of cellulose. *FEMS Microbiology Letters*, **13**, 25–58.
13. Béguin, P. & Lemaire, M. (1996) The cellulosome: An exocellular, multiprotein complex specialized in cellulose degradation. *Critical Reviews in Biochemistry and Molecular Biology*, **31**, 201–236.
14. Warren, R.A.J. (1996) Microbial hydrolysis of polysaccharides. *Annual Review of Microbiology*, **50**, 183–212.
15. Teeri, T.T., Koivula, A., Linder, M., Wohlfahrt, G., Divne, C. & Jones, T.A. (1998) *Trichoderma reesei* cellobiohydrolases: Why so efficient on crystalline cellulose? *Biochemical Society Transactions*, **26**, 173–178.
16. Lynd, L.R., Weimer, P.J., van Zyl, W.H. & Pretorius, I.S. (2002) Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, **66**, 506–577.
17. Lowe, S.E., Jain, M.K. & Zeikus, J.G. (1993) Biology, ecology, and biotechnological applications of anaerobic bacteria adapted to environmental stresses in temperature, pH, salinity, or substrates. *Microbiological Reviews*, **57**, 451–509.
18. Lamed, R. & Bayer, E.A. (1988) The cellulosome of *Clostridium thermocellum*. *Advances in Applied Microbiology*, **33**, 1–46.
19. Bayer, E.A., Morag, E. & Lamed, R. (1994) The cellulosome – A treasure-trove for biotechnology. *Trends in Biotechnology*, **12**, 378–386.
20. Bayer, E.A., Shimon, L.J.W., Lamed, R. & Shoham, Y. (1998) Cellulosomes: Structure and ultrastructure. *Journal of Structural Biology*, **124**, 221–234.
21. Bayer, E.A., Shoham, Y. & Lamed, R. (2000) Cellulose-decomposing prokaryotes and their enzyme systems. In: *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community*, 3rd edn (latest update release 3.7, September 2001) (eds M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt). Springer-Verlag, New York. Available from: <http://link.springer.de/link/service/books/10125/index.htm>.

22. Bayer, E.A., Belaich, J.-P., Shoham, Y. & Lamed, R. (2004) The cellulosomes: Multi-enzyme machines for degradation of plant cell wall polysaccharides. *Annual Review of Microbiology*, **58**, 521–554.
23. Doi, R.H. & Kosugi, A. (2004) Cellulosomes: Plant-cell-wall-degrading enzyme complexes. *Nature Reviews Microbiology*, **2**, 541–551.
24. Demain, A.L., Newcomb, M. & Wu, J.H. (2005) Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews*, **69**, 124–154.
25. Felix, C.R. & Ljungdahl, L.G. (1993) The cellulosome – the exocellular organelle of *Clostridium*. *Annual Review of Microbiology*, **47**, 791–819.
26. Schwarz, W.H. (2001) The cellulosome and cellulose degradation by anaerobic bacteria. *Applied Microbiology and Biotechnology*, **56**, 634–649.
27. Bayer, E.A., Setter, E. & Lamed, R. (1985) Organization and distribution of the cellulosome in *Clostridium thermocellum*. *Journal of Bacteriology*, **163**, 552–559.
28. Bayer, E.A. & Lamed, R. (1986). Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *Journal of Bacteriology*, **167**, 828–836.
29. Lamed, E. & Bayer, E.A. (1986) Contact and cellulolysis in *Clostridium thermocellum* via extensile surface organelles. *Experientia*, **42**, 72–73.
30. Lamed, R., Naimark, J., Morgenstern, E. & Bayer, E.A. (1987) Scanning electron microscopic delineation of bacterial surface topology using cationized ferritin. *Journal of Microbiological Methods*, **7**, 233–240.
31. Anderson, K.L. (1998) Cationized ferritin as a stain for electron microscopic observation of bacterial ultrastructure. *Biotechnic and Histochemistry*, **73**, 278–288.
32. Mayer, F., Coughlan, M.P., Mori, Y. & Ljungdahl, L.G. (1987) Macromolecular organization of the cellulolytic enzyme complex of *Clostridium thermocellum* as revealed by electron microscopy. *Applied and Environmental Microbiology*, **53**, 2785–2792.
33. Mayer, F. (1988) Cellulolysis: Ultrastructural aspects of bacterial systems. *Electron Microscopy Reviews*, **1**, 69–85.
34. Madkour, M. & Mayer, F. (2003) Structural organization of the intact bacterial cellulosome as revealed by electron microscopy. *Cell Biology International*, **27**, 831–836.
35. Gerngross, U.T., Romaniec, M.P.M., Kobayashi, T., Huskisson, N.S. & Demain, A.L. (1993) Sequencing of a *Clostridium thermocellum* gene (cipA) encoding the cellulosomal S_L-protein reveals an unusual degree of internal homology. *Molecular Microbiology*, **8**, 325–334.
36. Fujino, T., Béguin, P. & Aubert, J.-P. (1993) Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. *Journal of Bacteriology*, **175**, 1891–1899.
37. Shoseyov, O., Takagi, M., Goldstein, M.A. & Doi, R.H. (1992) Primary sequence analysis of *Clostridium cellulovorans* cellulose binding protein A. *Proceedings of the National Academy of Sciences USA*, **89**, 3483–3487.
38. Poole, D.M., Morag, E., Lamed, R., Bayer, E.A., Hazlewood, G.P. & Gilbert, H.J. (1992) Identification of the cellulose binding domain of the cellulosome subunit S1 from *Clostridium thermocellum*. *FEMS Microbiology Letters*, **99**, 181–186.
39. Boraston, A.B., Bolam, D.N., Gilbert, H.J. & Davies, G.J. (2004) Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochemical Journal*, **382**, 769–781.
40. Leibovitz, E. & Béguin, P. (1996) A new type of cohesin domain that specifically binds the dockerin domain of the *Clostridium thermocellum* cellulosome-integrating protein CipA. *Journal of Bacteriology*, **178**, 3077–3084.
41. Leibovitz, E., Ohayon, H., Gounon, P. & Béguin, P. (1997) Characterization and subcellular localization of the *Clostridium thermocellum* scaffoldin dockerin binding protein SdbA. *Journal of Bacteriology*, **179**, 2519–2523.

42. Lemaire, M., Ohayon, H., Gounon, P., Fujino, T. & Béguin, P. (1995) OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domains to components of the cell envelope. *Journal of Bacteriology*, **177**, 2451–2459.
43. Salamiou, S., Tokatlidis, K., Béguin, P. & Aubert, J.-P. (1992) Involvement of separate domains of the cellulosomal protein S1 of *Clostridium thermocellum* in binding to cellulose and in anchoring of catalytic subunits to the cellulosome. *FEBS Letters*, **304**, 89–92.
44. Lamed, R., Naimark, J., Morgenstern, E. & Bayer, E.A. (1987) Specialized cell surface structures in cellulolytic bacteria. *Journal of Bacteriology*, **169**, 3792–3800.
45. Kakiuchi, M., Isui, A., Suzuki, K., Fujino, T., Fujino, E., Kimura, T., Karita, S., Sakka, K. & Ohmiya, K. (1998) Cloning and DNA sequencing of the genes encoding *Clostridium josui* scaffolding protein CipA and cellulase CelD and identification of their gene products as major components of the cellulosome. *Journal of Bacteriology*, **180**, 4303–4308.
46. Pagès, S., Belaich, A., Fierobe, H.-P., Tardif, C., Gaudin, C. & Belaich, J.-P. (1999) Sequence analysis of scaffolding protein CipC and ORFXp, a new cohesin-containing protein in *Clostridium cellulolyticum*: Comparison of various cohesin domains and subcellular localization of ORFXp. *Journal of Bacteriology*, **181**, 1801–1810.
47. Adams, J.J., Pal, G., Yam, K., Spencer, H.L., Jia, Z. & Smith, S.P. (2005) Purification and crystallization of a trimodular complex comprising the type II cohesin-dockerin interaction from the cellulosome of *Clostridium thermocellum*. *Acta Crystallographica: Section F, Structural Biology and Crystallization Communications*, **61**, 46–48.
48. Adams, J.J., Webb, B.A., Spencer, H.L. & Smith, S.P. (2005) Structural characterization of type II dockerin module from the cellulosome of *Clostridium thermocellum*: Calcium-induced effects on conformation and target recognition. *Biochemistry*, **44**, 2173–2182.
49. Adams, J.J., Pal, G., Jia, Z. & Smith, S.P. (2006) Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proceedings of the National Academy of Sciences USA*, **103**, 305–310.
50. Adams, J.J., Jang, C.J., Spencer, H.L., Elliott, M. & Smith, S.P. (2004) Expression, purification and structural characterization of the scaffoldin hydrophilic X-module from the cellulosome of *Clostridium thermocellum*. *Protein Expression and Purification*, **38**, 258–263.
51. Carvalho, A.L., Pires, V.M., Gloster, T.M., Turkenburg, J.P., Prates, J.A., Ferreira, L.M., Romão, M.J., Davies, G.J., Fontes, C.M. & Gilbert, H.J. (2005) Insights into the structural determinants of cohesin-dockerin specificity revealed by the crystal structure of the type II cohesin from *Clostridium thermocellum* SdbA. *Journal of Molecular Biology*, **349**, 909–915.
52. Smith, S.P., Béguin, P., Alzari, P.M. & Gehring, K. (2002) ^1H , ^{13}C , ^{15}N NMR sequence-specific resonance assignment of a *Clostridium thermocellum* type II cohesin module. *Journal of Biomolecular NMR*, **23**, 73–74.
53. Dror, T.W., Morag, E., Rolider, A., Bayer, E.A., Lamed, R. & Shoham, Y. (2003) Regulation of the cellulosomal *celS* (*cel48A*) gene of *Clostridium thermocellum* is growth-rate dependent. *Journal of Bacteriology*, **185**, 3042–3048.
54. Dror, T.W., Rolider, A., Bayer, E.A., Lamed, R. & Shoham, Y. (2003) Regulation of expression of scaffoldin-related genes in *Clostridium thermocellum*. *Journal of Bacteriology*, **185**, 5109–5116.
55. Dror, T.W., Rolider, A., Bayer, E.A., Lamed, R. & Shoham, Y. (2005) Regulation of major cellulosomal endoglucanases of *Clostridium thermocellum* differs from that of a prominent cellulosomal xylanase. *Journal of Bacteriology*, **187**, 2261–2266.
56. Miron, J., Yokoyama, M.T. & Lamed, R. (1989) Bacterial cell surface structures involved in lucerne cell wall degradation by pure cultures of cellulolytic rumen bacteria. *Applied Microbiology and Biotechnology*, **32**, 218–222.
57. Pohlschröder, M., Canale-Parola, E. & Leschine, S.B. (1995) Ultrastructural diversity of the cellulase complexes of *Clostridium papyrosolvens* C7. *Journal of Bacteriology*, **177**, 6625–6629.

58. Pohlschröder, M., Leschine, S.B. & Canale-Parola, E. (1993) Regulation of the multicomplex cellulase-xylanase system of *Clostridium papyrosolvens*. In: *Genetics, Biochemistry and Ecology of Lignocellulose Degradation* (eds. K. Shimada, S. Hoshino, K. Ohmiya, K. Sakka, Y. Kobayashi & S. Karita), pp. 86–94. Uni Publishers, Tokyo, Japan.
59. Pohlschröder, M., Leschine, S.B. & Canale-Parola, E. (1994) Multicomplex cellulase-xylanase system of *Clostridium papyrosolvens* C7. *Journal of Bacteriology*, **176**, 70–76.
60. Han, S.O., Yukawa, H., Inui, M. & Doi, R.H. (2005) Effect of carbon source on the cellulosomal subpopulations of *Clostridium cellulovorans*. *Microbiology*, **151**, 1491–1497.
61. Han, S.O., Cho, H.Y., Yukawa, H., Inui, M. & Doi, R.H. (2004) Regulation of expression of cellosomes and noncellulosomal (hemi)cellulolytic enzymes in *Clostridium cellulovorans* during growth on different carbon sources. *Journal of Bacteriology*, **186**, 4218–4227.
62. Han, S.O., Yukawa, H., Inui, M. & Doi, R.H. (2003) Regulation of expression of cellulosomal cellulase and hemicellulase genes in *Clostridium cellulovorans*. *Journal of Bacteriology*, **185**, 6067–6075.
63. Rincon, M.T., McCrae, S.I., Kirby, J., Scott, K.P. & Flint, H.J. (2001) EndB, a multidomain family 44 cellulase from *Ruminococcus flavefaciens* 17, binds to cellulose via a novel cellulose-binding module and to another *R. flavefaciens* protein via a dockerin domain. *Applied and Environmental Microbiology*, **67**, 4426–4431.
64. Aurilia, V., Martin, J.C., McCrae, S.I., Scott, K.P., Rincon, M.T. & Flint, H.J. (2000) Three multidomain esterases from the cellulolytic rumen anaerobe *Ruminococcus flavefaciens* 17 that carry divergent dockerin sequences. *Microbiology*, **146**, 1391–1397.
65. Rincon, M.T., Martin, J.C., Aurilia, V., McCrae, S.I., Rucklidge, G., Reid, M., Bayer, E.A., Lamed, R. & Flint, H.J. (2004) ScaC, an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the *Ruminococcus flavefaciens* 17 cellosome. *Journal of Bacteriology*, **186**, 2576–2585.
66. Ding, S.-Y., Rincon, M.T., Lamed, R., Martin, J.C., McCrae, S.I., Aurilia, V., Shoham, Y., Bayer, E.A. & Flint, H.J. (2001) Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *Journal of Bacteriology*, **183**, 1945–1953.
67. Rincon, M.T., Ding, S.-Y., McCrae, S.I., Martin, J.C., Aurilia, V., Lamed, R., Shoham, Y., Bayer, E.A. & Flint, H.J. (2003) Novel organization and divergent dockerin specificities in the cellosome system of *Ruminococcus flavefaciens*. *Journal of Bacteriology*, **185**, 703–713.
68. Rincon, M.T., Cepeljnik, T., Martin, J.C., Lamed, R., Barak, Y., Bayer, E.A. & Flint, H.J. (2005) Unconventional mode of attachment of the *Ruminococcus flavefaciens* cellosome to the cell surface. *Journal of Bacteriology*, **187**, 7569–7578.
69. Navarre, W.W. & Schneewind, O. (1994) Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. *Molecular Microbiology*, **14**, 115–121.
70. Navarre, W.W. & Schneewind, O. (1999) Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiology and Molecular Biology Reviews*, **63**, 174–229.
71. Ton-That, H., Marraffini, L.A. & Schneewind, O. (2004) Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochimica et Biophysica Acta*, **1694**, 269–278.
72. Schneewind, O., Fowler, A. & Faull, K.F. (1995) Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science*, **268**, 103–106.
73. Grépinet, O., Chebrou, M.-C. & Béguin, P. (1988) Nucleotide sequence and deletion analysis of the xylanase gene (*xynZ*) of *Clostridium thermocellum*. *Journal of Bacteriology*, **170**, 4582–4588.
74. Hazlewood, G.P., Romaniec, M.P.M., Davidson, K., Grépinet, O., Béguin, P., Millet, J., Raynaud, O. & Aubert, J.-P. (1988) A catalogue of *Clostridium thermocellum* endoglucanase, b-glucosidase and xylanase genes cloned in *Escherichia coli*. *FEMS Microbiology Letters*, **51**, 231–236.

75. Blum, D.L., Kataeva, I.A., Li, X.L. & Ljungdahl, L.G. (2000) Feruloyl esterase activity of the *Clostridium thermocellum* cellulosome can be attributed to previously unknown domains of XynY and XynZ. *Journal of Bacteriology*, **182**, 1346–1351.
76. Morag, E., Bayer, E.A. & Lamed, R. (1990) Relationship of cellulosomal and noncellulosomal xylanases of *Clostridium thermocellum* to cellulose-degrading enzymes. *Journal of Bacteriology*, **172**, 6098–6105.
77. Zverlov, V.V., Fuchs, K.-P. & Schwarz, W.H. (2002) Chi18A, the endochitinase in the cellulosome of the thermophilic, cellulolytic bacterium *Clostridium thermocellum*. *Applied and Environmental Microbiology*, **68**, 3176–3179.
78. Zverlov, V.V., Fuchs, K.P., Schwarz, W.H. & Velikodvorskaya, G. (1994) Purification and cellulosomal localization of *Clostridium thermocellum* mixed linkage b-glucanase LicB (1,3-1,4-b-D-glucanase). *Biotechnology Letters*, **16**, 29–34.
79. Halstead, J.R., Vercoc, P.E., Gilbert, H.J., Davidson, K. & Hazlewood, G.P. (1999) A family 26 mannanase produced by *Clostridium thermocellum* as a component of the cellulosome contains a domain which is conserved in mannanases from anaerobic fungi. *Microbiology*, **45**, 3101–3108.
80. Tamaru, Y. & Doi, R.H. (2001) Pectate lyase A, an enzymatic subunit of the *Clostridium cellulovorans* cellulosome. *Proceedings of the National Academy of Sciences USA*, **20**, 4125–4129.
81. Pagès, S., Valette, O., Abdou, L., Belaich, A. & Belaich, J.-P. (2003) A rhamnogalacturonan lyase in the *Clostridium cellulolyticum* cellulosome. *Journal of Bacteriology*, **185**, 4727–4733.
82. Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I., Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M.J., Bennett, G.N., Koonin, E.V. & Smith, D.R. (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *Journal of Bacteriology*, **183**, 4823–4838.
83. Sabathe, F., Belaich, A. & Soucaille, P. (2002) Characterization of the cellulolytic complex (cellulosome) of *Clostridium acetobutylicum*. *FEMS Microbiology Letters*, **217**, 15–22.
84. Sabathe, F. & Soucaille, P. (2003) Characterization of the CipA scaffolding protein and in vivo production of a minicellulosome in *Clostridium acetobutylicum*. *Journal of Bacteriology*, **185**, 1092–1096.
85. Antonopoulos, D.A., Nelson, K.E., Morrison, M. & White, B.A. (2004) Strain-specific genomic regions of *Ruminococcus flavefaciens* FD-1 as revealed by combinatorial random-phase genome sequencing and suppressive subtractive hybridization. *Environmental Microbiology*, **6**, 335–346.
86. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*, **280**, 309–316.
87. Henrissat, B. & Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal*, **293**, 781–788.
88. Lamed, R., Setter, E. & Bayer, E.A. (1983) Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *Journal of Bacteriology*, **156**, 828–836.
89. Morag, E., Halevy, I., Bayer, E.A. & Lamed, R. (1991) Isolation and properties of a major cellobiohydrolase from the cellulosome of *Clostridium thermocellum*. *Journal of Bacteriology*, **173**, 4155–4162.
90. Wu, J.H.D., Orme-Johnson, W.H. & Demain, A.L. (1988) Two components of an extracellular protein aggregate of *Clostridium thermocellum* together degrade crystalline cellulose. *Biochemistry*, **27**, 1703–1709.
91. Ali, B.R., Romaniec, M.P., Hazlewood, G.P. & Freedman, R.B. (1995) Characterization of the subunits in an apparently homogeneous subpopulation of *Clostridium thermocellum* cellulosomes. *Enzyme and Microbial Technology*, **17**, 705–711.
92. Wang, W.K., Kruus, K. & Wu, J.H.D. (1993) Cloning and DNA sequence of the gene coding for *Clostridium thermocellum* cellulase S_S (CelS), a major cellulosome component. *Journal of Bacteriology*, **175**, 1293–1302.

93. Zverlov, V.V., Kellermann, J. & Schwarz, W.H. (2005) Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: Identification of the major catalytic components in the extra-cellular complex and detection of three new enzymes. *Proteomics*, **5**, 3646–3653.
94. Bayer, E.A., Shoham, Y. & Lamed, R. (2000) The cellulosome – an exocellular organelle for degrading plant cell wall polysaccharides. In: *Glycomicrobiology* (ed. R.J. Doyle), pp. 387–439. Kluwer Academic/Plenum Publishers, New York.
95. Schwarz, W.H., Zverlov, V.V. & Bahl, H. (2004) Extracellular glycosyl hydrolases from Clostridia. *Advances in Applied Microbiology*, **56**, 215–261.
96. Ahsan, M.M., Kimura, T., Karita, S., Sakka, K. & Ohmiya, K. (1996) Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. *Journal of Bacteriology*, **178**, 5732–5740.
97. Ahsan, M.M., Matsumoto, M., Karita, S., Kimura, T., Sakka, K. & Ohmiya, K. (1997) Purification and characterization of the family J catalytic domain derived from the *Clostridium thermocellum* endoglucanase CelJ. *Bioscience, Biotechnology, and Biochemistry*, **61**, 427–431.
98. Zverlov, V.V., Velikodvorskaya, G.V., Schwarz, W.H., Bronnenmeier, K., Kellermann, J. & Staudenbauer, W.L. (1998) Multidomain structure and cellulosomal localization of the *Clostridium thermocellum* cellobiohydrolase CbhA. *Journal of Bacteriology*, **180**, 3091–3099.
99. Fernandes, A.C., Fontes, C.M., Gilbert, H.J., Hazlewood, G.P., Fernandes, T.H. & Ferreira, L.M.A. (1999) Homologous xylanases from *Clostridium thermocellum*: Evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. *Biochemical Journal*, **342**, 105–110.
100. Prates, J.A., Tarbouriech, N., Charnock, S.J., Fontes, C.M., Ferreira, L.M. & Davies, G.J. (2001) The structure of the feruloyl esterase module of xylanase 10B from *Clostridium thermocellum* provides insights into substrate recognition. *Structure (Cambridge)*, **9**, 1183–1190.
101. Tarbouriech, N., Prates, J.A., Fontes, C.M. & Davies, G.J. (2005) Molecular determinants of substrate specificity in the feruloyl esterase module of xylanase 10B from *Clostridium thermocellum*. *Acta Crystallographica, Section D. Biological Structure*, **61**, 194–197.
102. Linder, M. & Teeri, T.T. (1997) The roles and function of cellulose-binding domains. *Journal of Biotechnology*, **57**, 15–28.
103. Tomme, P., Warren, R.A.J., Miller, R.C., Kilburn, D.G. & Gilkes, N.R. (1995) Cellulose-binding domains – Classification and properties. In: *Enzymatic Degradation of Insoluble Polysaccharides* (eds J.M. Saddler & M.H. Penner), pp. 142–161. American Chemical Society, Washington, DC.
104. Carrard, G., Koivula, A., Soderlund, H. & Béguin, P. (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proceedings of the National Academy of Sciences USA*, **97**, 10342–10347.
105. Blake, A.W., McCartney, L., Flint, J.E., Bolam, D.N., Boraston, A.B., Gilbert, H.J. & Knox, J.P. (2006) Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *Journal of Biological Chemistry*, **281**, 29321–29329.
106. Bayer, E.A., Coutinho, P.M. & Henrissat, B. (1999) Cellulosome-like sequences in *Archaeoglobus fulgidus*: An enigmatic vestige of cohesin and dockerin domains. *FEBS Letters*, **463**, 277–280.
107. Xie, G., Bruce, D.C., Challacombe, J.F., Chertkov, O., Detter, J.C., Gilna, P., Han, C.S., Lucas, S., Misra, M., Myers, G.L., Richardson, P., Tapia, R., Thayer, N., Thompson, L.S., Brettin, T.S., Henrissat, B., Wilson, D.B. & McBride, M.J. (2007) Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. *Applied and Environmental Microbiology*, **73**, 3536–3546.
108. Bayer, E.A., Kenig, R. & Lamed, R. (1983) Adherence of *Clostridium thermocellum* to cellulose. *Journal of Bacteriology*, **156**, 818–827.
109. Bayer, E.A., Morag, E., Shoham, Y., Tormo, J. & Lamed, R. (1996) The cellulosome: A cell-surface organelle for the adhesion to and degradation of cellulose. In: *Bacterial Adhesion: Molecular and Ecological Diversity* (ed. M. Fletcher), pp. 155–182. Wiley-Liss, Inc, New York.

110. Morag, E., Lapidot, A., Govorko, D., Lamed, R., Wilchek, M., Bayer, E.A. & Shoham, Y. (1995) Expression, purification and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of *Clostridium thermocellum*. *Applied and Environmental Microbiology*, **61**, 1980–1986.
111. Lamed, R., Kenig, R., Setter, E. & Bayer, E.A. (1985) Major characteristics of the cellulolytic system of *Clostridium thermocellum* coincide with those of the purified cellulosome. *Enzyme and Microbial Technology*, **7**, 37–41.
112. Tormo, J., Lamed, R., Chirino, A.J., Morag, E., Bayer, E.A., Shoham, Y. & Steitz, T.A. (1996) Crystal structure of a bacterial family-III cellulose-binding domain: A general mechanism for attachment to cellulose. *EMBO Journal*, **15**, 5739–5751.
113. Jindou, S., Xu, Q., Kenig, R., Shoham, Y., Bayer, E.A. & Lamed, R. (2006) Novel architectural theme of family-9 glycoside hydrolases identified in cellulosomal enzymes of *Acetivibrio cellulolyticus* and *Clostridium thermocellum*. *FEMS Microbiology Letters*, **254**, 308–316.
114. Gilad, R., Rabinovich, L., Yaron, S., Bayer, E.A., Lamed, R., Gilbert, H.J. & Shoham, Y. (2003) Cell, a non-cellulosomal family-9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. *Journal of Bacteriology*, **185**, 391–398.
115. Berger, E., Zhang, D., Zverlov, V.V. & Schwarz, W.H. (2007) Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. *FEMS Microbiology Letters*, **268**, 194–201.
116. Sakon, J., Irwin, D., Wilson, D.B. & Karplus, P.A. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nature Structural Biology*, **4**, 810–818.
117. Irwin, D., Shin, D.-H., Zhang, S., Barr, B.K., Sakon, J., Karplus, P.A. & Wilson, D.B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *Journal of Bacteriology*, **180**, 1709–1714.
118. Kataeva, I.A., Seidel, R.D., III, Shah, A., West, L.T., Li, X.-L. & Ljungdahl, L.G. (2002) The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. *Applied and Environmental Microbiology*, **68**, 4292–4300.
119. Zverlov, V.V., Velikodvorskaya, G.A. & Schwarz, W.H. (2002) A newly described cellulosomal cellobiohydrolase, CelO, from *Clostridium thermocellum*: Investigation of the exo-mode of hydrolysis, and binding capacity to crystalline cellulose. *Microbiology*, **148**, 247–255.
120. Lamed, R., Setter, E., Kenig, R. & Bayer, E.A. (1983) The cellulosome – a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. *Biotechnology and Bioengineering Symposium*, **13**, 163–181.
121. Lemaire, M., Miras, I., Gounon, P. & Béguin, P. (1998) Identification of a region responsible for binding to the cell wall within the S-layer protein of *Clostridium thermocellum*. *Microbiology*, **144**, 211–217.
122. Xu, Q., Gao, W., Ding, S.-Y., Kenig, R., Shoham, Y., Bayer, E.A. & Lamed, R. (2003) The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell-surface anchoring protein. *Journal of Bacteriology*, **185**, 4548–4557.
123. Xu, Q., Barak, Y., Kenig, R., Shoham, Y., Bayer, E.A. & Lamed, R. (2004) A novel *Acetivibrio cellulolyticus* anchoring scaffoldin that bears divergent cohesins. *Journal of Bacteriology*, **186**, 5782–5789.
124. Xu, Q., Bayer, E.A., Goldman, M., Kenig, R., Shoham, Y. & Lamed, R. (2004) Architecture of the *Bacteroides cellulosolvens* cellulosome: Description of a cell-surface anchoring scaffoldin and a family-48 cellulase. *Journal of Bacteriology*, **186**, 968–977.
125. Doi, R.H., Kosugi, A., Murashima, K., Tamaru, Y. & Han, S.O. (2003) Cellulosomes from mesophilic bacteria. *Journal of Bacteriology*, **185**, 5907–5914.
126. Belaich, J.-P., Belaich, A., Fierobe, H.-P., Gal, L., Gaudin, C., Pagès, S., Reverbel-Leroy, C. & Tardif, C. (1999) The cellulolytic system of *Clostridium cellulolyticum*. In: *Genetics, Biochemistry*

- and *Ecology of Cellulose Degradation* (eds. K. Ohmiya, K. Hayashi, K. Sakka, Y. Kobayashi, S. Karita & T. Kimura), pp. 479–487. Uni Publishers, Tokyo.
127. Kosugi, A., Amano, Y., Murashima, K. & Doi, R.H. (2004) Hydrophilic domains of scaffolding protein CbpA promote glycosyl hydrolase activity and localization of cellulosomes to the cell surface of *Clostridium cellulovorans*. *Journal of Bacteriology*, **186**, 6351–6359.
 128. Bayer, E.A. & Lamed, R. (1992) The cellulose paradox: Pollutant *par excellence* and/or a reclaimable natural resource? *Biodegradation*, **3**, 171–188.
 129. Lamed, R. & Bayer, E.A. (1988) The cellulosome concept: Exocellular/extracellular enzyme reactor centers for efficient binding and cellulolysis. In: *Biochemistry and Genetics of Cellulose Degradation* (eds. J.-P. Aubert, P. Beguin & J. Millet), pp. 101–116. Academic Press, London.
 130. Lamed, R. & Bayer, E.A. (1991) Cellulose degradation by thermophilic anaerobic bacteria. In: *Biosynthesis and Biodegradation of Cellulose and Cellulose Materials* (eds. C.H. Haigler & P.J. Weimer), pp. 377–410. Marcel Dekker, New York.
 131. Kulshreshtha, A.K. & Dweltz, N.E. (1973) Paracrystalline lattice disorder in cellulose. I. Reappraisal of the application of the two-phase hypothesis to the analysis of powder x-ray diffractograms of native and hydrolyzed cellulosic materials. *Journal of Polymer Science: Polymer Physics Edition*, **11**, 487–497.
 132. Wood, T.M. (1988) Preparation of crystalline, amorphous and dyed cellulase substrates. *Methods in Enzymology*, **160**, 19–25.
 133. Boisset, C., Chanzy, H., Henrissat, B., Lamed, R., Shoham, Y. & Bayer, E.A. (1999) Digestion of crystalline cellulose substrates by the *Clostridium thermocellum* cellulosome: Structural and morphological aspects. *Biochemical Journal*, **340**, 829–835.
 134. Chanzy, H. & Henrissat, B. (1983) Electron microscopy study of the enzymic hydrolysis of *Valonia* cellulose. *Carbohydrate Polymers*, **3**, 161–173.
 135. Chanzy, H. & Henrissat, B. (1985) Unidirectional degradation of *Valonia* cellulose microcrystals subjected to cellulase action. *FEBS Letters*, **184**, 285–288.
 136. Chanzy, H., Henrissat, B., Vuong, R. & Schülein, M. (1983) The action of 1,4-b-D-glucan cellobiohydrolase on *Valonia* cellulose microcrystals. An electron microscopic study. *FEBS Letters*, **153**, 113–118.
 137. Boisset, C., Fraschini, C., Schulein, M., Henrissat, B. & Chanzy, H. (2000) Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from *Humicola insolens* and its mode of synergy with cellobiohydrolase Cel7A. *Applied and Environmental Microbiology*, **66**, 1444–1452.
 138. Imai, T., Boisset, C., Samejima, M., Igarashi, K. & Sugiyama, J. (1998) Unidirectional processive action of cellobiohydrolase Cel7A on *Valonia* cellulose microcrystals. *FEBS Letters*, **432**, 113–116.
 139. Parsiegla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Belaich, J.P., Driguez, H. & Haser, R. (1998) The crystal structure of the processive endocellulase CelF of *Clostridium cellulolyticum* in complex with a thiooligosaccharide inhibitor at 2.0 Å resolution. *EMBO Journal*, **17**, 5551–5562.
 140. Johnson, E.A., Sakojoh, M., Halliwell, G., Madia, A. & Demain, A.L. (1982) Saccharification of complex cellulosic substrates by the cellulase system from *Clostridium thermocellum*. *Applied and Environmental Microbiology*, **43**, 1125–1132.
 141. Zhang, Y. & Lynd, L.R. (2003) Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: Development of an enzyme-linked immunosorbent assay-based method with application to *Clostridium thermocellum* batch cultures. *Analytical Chemistry*, **75**, 219–227.
 142. Morag, E., Yaron, S., Lamed, R., Kenig, R., Shoham, Y. & Bayer, E.A. (1996) Dissociation of the cellulosome of *Clostridium thermocellum* under nondenaturing conditions. *Journal of Biotechnology*, **51**, 235–242.

143. Lamed, R., Kenig, R., Morag (Morgenstern), E., Calzada, J.F., de Micheo, F. & Bayer, E.A. (1991) Efficient cellulose solubilization by a combined cellulosome-b-glucosidase system. *Applied Biochemistry and Biotechnology*, **27**, 173–183.
144. Fierobe, H.-P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J.-P. & Bayer, E.A. (2001) Design and production of active cellulosome chimeras: Selective incorporation of dockerin-containing enzymes into defined functional complexes. *Journal of Biological Chemistry*, **276**, 21257–21261.
145. Fierobe, H.-P., Bayer, E.A., Tardif, C., Czjzek, M., Mechaly, A., Belaich, A., Lamed, R., Shoham, Y. & Belaich, J.-P. (2002) Degradation of cellulose substrates by cellulosome chimeras: Substrate targeting versus proximity of enzyme components. *Journal of Biological Chemistry*, **277**, 49621–49630.
146. Bayer, E.A., Lamed, R. & Himmel, M.E. (2007) The potential of cellulases and cellulosomes for cellulosic waste management. *Current Opinion in Biotechnology*, **18**, 237–245.

Chapter 14

Pretreatments for Enhanced Digestibility of Feedstocks

David K. Johnson, and Richard T. Elander

This chapter reviews how pretreatment is used to decrease the recalcitrance of biomass, and make the cellulose in the feedstock more susceptible to digestion by cellulase enzymes. Pretreatments have been designed to operate under very different chemistries, at widely varying temperatures, and for very different reaction times. With some pretreatments there is almost no change in the composition of the feedstock whereas in others the composition is changed considerably by dissolving the hemicelluloses, the lignin, or both. Pretreatment can also cause changes in the physicochemical structure of the biomass with changes in cellulose crystallinity, molecular weight, and accessibility, plus changes in biomass porosity and particle size. This chapter will focus on more recent developments in pretreatment and describe how these pretreatment processes attempt to overcome the natural recalcitrance of biomass to digestion by enzymes.

14.1 Introduction

Prior to the discovery of cellulolytic enzyme systems, processes that could thermochemically hydrolyze lignocellulose to produce soluble sugars were investigated and developed with some commercial use during wartime periods. Most of these processes were operated under fairly severe conditions and typically utilized concentrated acids (primarily sulfuric or hydrochloric) at relatively low temperatures (under 100°C) or dilute acids (again, typically sulfuric or hydrochloric) at much higher temperatures (often above 200°C). Such harsh conditions generally resulted in relatively low recoverable sugar yields due to sugar degradation reactions that produced primarily aldehydes and organic acids, along with other undesirable compounds. Also, the harsh conditions and, in the case of concentrated acid processes, the large amounts of catalyst used, caused such processes to be highly capital intensive, either from the requirements for pressurized corrosion resistant reactors capable of processing solid materials or the economic requirement to recover and recycle the acid catalysts. Advances have been made in thermochemical acid hydrolysis processes to improve their commercialization potential in some niche opportunities, especially for concentrated acid processes. These improvements are primarily related to acid catalyst recovery processes that allow for the cost-effective recycling of the large quantities of required acid catalysts (1).

The discovery of cellulase enzymes and the commercial availability of such enzyme systems have dramatically changed the context of the thermochemical step in the conversion of lignocellulosic biomass to sugars. Rather than requiring this thermochemical step to directly produce all resulting sugars from biomass, this step can now be viewed as a true pretreatment step, whose purpose is to prepare the biomass for rapid and complete enzymatic hydrolysis to produce a monomeric sugar stream. Conceptually, the enzymatic hydrolysis approach has several inherent advantages over thermochemical hydrolysis, including low-temperature, mild-pH conditions leading to less expensive and less complex reactor systems and significantly lower or no losses of resulting sugars to degradation products.

14.2 Enzyme usage and enzyme-type considerations for pretreated biomass

Different pretreatment approaches catalyze biomass hydrolysis and other reactions to various extents. Therefore, the composition of the liquid and solid process streams resulting from different pretreatment approaches can be widely different. For example, dilute sulfuric acid pretreatment processes can solubilize nearly all the hemicellulose but very little lignin or cellulose. Other pretreatment approaches, primarily the alkaline processes, are more effective at solubilizing lignin, but leave extensive amounts of the hemicellulose fraction as an insoluble component of the pretreated solids. These factors greatly impact the relative composition of the pretreated solids and the requirements for effective enzymatic saccharification in subsequent processing steps.

While there are considerable economic drivers to reduce the overall severity of the pretreatment operation (lower reactor materials cost, lower temperature and/or residence time, lower losses of resulting sugars to degradation, lower requirements to adequately “condition” pretreatment hydrolyzates for subsequent fermentation processes), less aggressive pretreatment conditions will generally result in less sugar release (primarily from hemicellulose) in the pretreatment step. This will shift more of the hydrolytic sugar production requirement from the pretreatment step to the enzymatic hydrolysis step and will have an impact on the amount and type of enzymes required to achieve high sugar yields from both cellulose and hemicellulose in less severely pretreated biomass.

Recently, greater attention is being given to the understanding and development of hemicellulase and other “accessory” enzyme systems needed to effectively debranch and depolymerize residual insoluble hemicellulose and soluble hemicellulose oligomers resulting from less severely pretreated biomass processes (2). Systematic studies are needed to determine whether augmentation or partial substitution of cellulase activity with various hemicellulase and accessory enzyme activities can improve the cost-effectiveness of biomass conversion processes by increasing sugar yields.

14.3 Desired properties of pretreatment processes

While no one particular pretreatment process can presently be viewed as the “ideal” approach for all feedstocks or for all process circumstances, a well-accepted list of the desired

properties of an ideal pretreatment process has been generated (3). Such an ideal pretreatment process:

- Produces a highly digestible pretreated solid
- Does not significantly degrade pentoses
- Does not significantly inhibit subsequent fermentation steps
- Requires little or no size reduction of biomass feedstock
- Can work in reactors of reasonable size and moderate cost
- Produces no solid-waste residues
- Has a high degree of simplicity
- Is effective at low moisture content

14.4 Physicochemical properties of pretreated biomass believed to affect cellulose digestibility

A significant effort has been expended, aimed at increasing our understanding of the factors that are the most critical in controlling the susceptibility of cellulosic substrates to enzymatic hydrolysis. A wide variety of physical and chemical properties of pretreated biomass and the cellulose remaining after pretreatment have been studied and reviewed (4–7). Changes in the physicochemical properties of pretreated lignocellulosics have been correlated with the enzymatic hydrolysis of the cellulose. There are general conclusions that have come out of these reviews that should be considered when trying to identify the critical properties in pretreated biomass. As stated by Coughlan (4), “There is considerable disagreement in the literature regarding the relative importance of the various factors that render cellulose so recalcitrant to hydrolysis.” Mansfield *et al.* (5) cautions that “when contemplating these characteristics (structural and physicochemical features of the substrate) and identifying potential contributing factors or limitations (to enzymatic hydrolysis) care must be taken to consider some undisputable principles: (i) all samples of insoluble cellulose (both native and pretreated) are structurally non-uniform; (ii) the pretreatment method and conditions can effectively alter the structure of the original cellulose; (iii) native cellulose contains inherent regions of highly ordered and disordered molecular polymers (i.e., crystalline and amorphous regions); (iv) considerable attention must be paid to the anatomical and structural ‘levels’ of organization (i.e., microfibril, fibril, or fiber) which is being modified or characterized during hydrolysis...”

A key factor for successful enzymatic conversion of biomass to fermentable sugars is the accessibility of the $\beta(1\rightarrow4)$ glycosidic bonds in cellulose to cellulase enzymes. Pretreatment regimes must be designed to remove substrate-specific barriers to cellulases to improve cellulose digestion. The precise nature of the obstacles encountered by cellulases in the complex biomass ultra-structure remains ambiguous. The effect of pretreatment is typically evaluated on the basis of improved enzyme digestibility and downstream ethanol production. The link between changes in cell wall chemistry/structure and cellulase digestibility is ultimately dependent on improved access to the cellulose microfibril. Accurate and direct assessment

Table 14.1 Effect of various pretreatment methods on the chemical composition and chemical/physical structure of lignocellulosic biomass (3)

	Increases accessible surface area	Decrystallizes cellulose	Removes hemicellulose	Removes lignin	Alters lignin structure
Uncatalyzed steam explosion	XX		XX		X
Liquid hot water	XX	ND	XX		X
pH Controlled hot water	XX	ND	XX		ND
Flow through liquid hot water	XX	ND	XX	X	X
Dilute acid	XX		XX		XX
Flow through acid	XX		XX	X	XX
AFEX	XX	XX	X	XX	XX
ARP	XX	XX	X	XX	XX
Lime	XX	ND	X	XX	XX

Note: XX, Strong effect; X, moderate effect; ND, not determined.

of changes in enzyme accessibility is challenging primarily due to the complexities of both the cellulase system and the biomass. (8).

Some of the factors that could influence the rate of enzymatic hydrolysis of cellulose in pretreated lignocellulosic feedstocks are cellulose crystallinity, degree of cellulose polymerization, feedstock particle size, the lignin barrier (content and distribution), substrate available surface area (pore volume), and cell wall thickness (coarseness). In addition, irreversible binding of enzymes onto lignin is also influenced by the nature of the substrate (6, 7). Typical physicochemical properties of biomass obtained from the pretreatments used by the CAFI group are shown in Table 14.1. Zhang and Lynd (9) have attempted to take this a step further by building functional models of cellulose hydrolysis that incorporate substrate features in addition to concentration and the activities of multiple cellulase components.

14.5 Pretreatment approaches

A large number of pretreatment approaches have been investigated across a variety of biomass feedstock types. Published studies are widely available and there are several review articles available that provide a general overview of the field (10–14). Unfortunately, standard experimental and analytical methodologies have not been utilized across much of the published pretreatment literature, making it difficult to conduct comparative evaluations based on published findings. Recently, several pretreatment research teams across North America have undertaken the first broad-ranging coordinated effort to develop comparative process performance and economic evaluation data for several leading pretreatment options. While this collaboration, known as the Biomass Refining Consortium for Applied

Fundamentals and Innovation (CAFI), does not fully encompass all possible pretreatment technologies or potential biomass feedstocks, it does serve as a model for how comparative data can be developed and made available for various stakeholders and potential commercializers of biomass conversion technologies. A series of papers that cover the comparative findings from a recently completed CAFI project on a common corn stover feedstock have been published (15–22).

14.5.1 Physical pretreatments

14.5.1.1 Comminution

Most pretreatment approaches require that collected biomass undergo some degree of mechanical size reduction prior to introduction into a pretreatment reactor. Woody biomass can be chipped in a manner similar to that commonly practiced in the pulp and paper industry. Depending on the pretreatment process and associated heat and mass transfer considerations, woody biomass is commonly comminuted to particle sizes smaller than typical wood pulping chips, as pretreatment processes are often practiced at much shorter residence times than wood pulping processes. Other biomass feedstocks, such as agricultural residues and herbaceous energy crops, can be coarsely chopped during or after the feedstock harvesting operation. Again, further comminution may be employed on these feedstock types as well, depending on the pretreatment process and associated heat and mass transfer considerations.

Intensive comminution of various biomass types has been practiced as an actual pretreatment process, without any further pretreatment prior to enzymatic hydrolysis. Methods include various types of ball milling (dry, wet, and vibratory processes), other types of attrition milling, compression milling, and wet or dry disk refining (23–27). While these methods can increase the enzymatic digestibility of the comminuted biomass by increasing the available surface area and by decrystallizing cellulose, most studies have concluded that the high mechanical power requirements cause comminution to be cost-prohibitive for use as a stand-alone pretreatment in a biomass to ethanol conversion process.

14.5.1.2 Irradiation

The use of high-energy electron beam and microwave energy sources as a biomass pretreatment approach has been investigated. These methods are believed to mechanically disrupt plant cell wall structure and decrease the crystallinity of cellulose, resulting in an increased enzymatic digestibility of cellulose. Issues with cost, energy intensity, and the practicality of applying such approaches in commercial processes have limited the development of irradiation as a viable pretreatment approach (12, 28).

14.5.2 Rapid decompression pretreatments

14.5.2.1 Steam explosion

Steam explosion processes with no added chemical catalysts have been practiced for nearly a century, dating back to the development of the Masonite process on wood chips in the

1920s (29). In steam explosion, chipped or coarsely shredded biomass is contacted with high-pressure saturated steam at high solids loadings (generally >20% solids) in a pressure vessel for a residence time that is generally 20 minutes or less (12–14, 30). Depending on the feedstock used and the objective of the pretreatment, steam explosion pretreatment temperatures are generally in the range of 140–260°C. At the end of the pretreatment time, the pressure vessel contents are rapidly decompressed into an atmospheric pressure flash tank, which causes significant disruption and defibration of the biomass. The use of specially designed orifices or other shear-enhancing devices to increase the mechanical disruption of the pretreated biomass is often practiced.

Even without the addition of any chemical catalysts, hydrolysis reactions in steam explosion are catalyzed by the release of organic acids that are liberated from acetyl functional groups associated with hemicellulose. This results in some lignin solubilization and hemicellulose hydrolysis, although yields of xylose from the hemicellulose fraction of most biomass types are typically no higher than 65% of theoretical, primarily due to extensive sugar degradation reactions that occur under typical uncatalyzed steam explosion reaction conditions (12, 31, 32). Compared to uncatalyzed steam explosion pretreatment, the addition of acid catalysts, such as sulfur dioxide or sulfuric acid, to biomass feedstocks in steam explosion, has been shown to improve the yield of released carbohydrates by reducing sugar degradation (13, 14, 33) and to improve the enzymatic digestibility of the resulting pretreated biomass (11, 13, 34). In general, steam explosion pretreatments have been shown to be effective in increasing the enzymatic digestibility of a wide range of feedstocks, although it is less effective on softwoods (35). This technique has also been tested in continuous pilot-scale reactor systems (36, 37), but it does result in some sugar degradation losses, incomplete disruption of the lignin–carbohydrate matrix, and the requirement to wash the pretreated biomass or otherwise condition the liquid portion of the pretreated slurry to remove inhibitory products prior to fermentation to ethanol or other products (11, 14, 33).

14.5.2.2 Ammonia fiber expansion (AFEX)

In many respects, the AFEX process is the alkaline equivalent of sulfur dioxide-catalyzed steam explosion pretreatment (13). In the AFEX process, biomass is treated with liquid anhydrous ammonia at temperatures between 60 and 100°C and pressures of 250–300 psig with residence times of about 5 minutes (20). The pressure is then released, resulting in a rapid expansion of the ammonia gas that causes swelling and physical disruption of biomass fibers and partial decrystallization of cellulose, along with some lignin solubilization and re-arrangement and some solubilization of hemicellulose primarily to oligomeric sugars (10, 14). AFEX is typically conducted at high solids loadings (about 40% solids) and high ammonia loadings (about 1.0 g NH₃/g dry feedstock), although the rapid expansion and high volatility of ammonia may permit near-complete recovery and recycle of ammonia (20, 38). The associated complexity and costs of ammonia recovery processes may be significant and must be better understood in order to assess the commercial potential of the AFEX process (21).

AFEX has been shown to deacetylate and increase the digestibility of biomass (39–41), although it does require that both cellulose and hemicellulose be enzymatically hydrolyzed due to limited hemicellulose hydrolysis during AFEX pretreatment. The AFEX

pretreatment is more effective on agricultural residues and herbaceous crops, with limited effectiveness demonstrated on woody biomass and other high-lignin feedstocks (14). AFEX has largely been practiced as a bench scale technique, although a larger, continuous version of AFEX based on extrusion technology, known as FIBEX, has been developed and tested (42).

14.5.3 Autohydrolysis pretreatments

14.5.3.1 Liquid hot water batch pretreatment

In addition to uncatalyzed steam explosion pretreatments, other uncatalyzed pretreatment processes using pressurized liquid hot water without rapid decompression have been investigated. Process conditions have been developed for cellulose hydrolysis at very high temperatures of about 260°C (10, 43) and as a pretreatment approach for achieving hemicellulose hydrolysis at lower temperatures of about 200–230°C (44, 45). High yields of soluble sugars from the hemicellulose fraction of some biomass types (primarily herbaceous crops and agricultural residues) can be achieved, but liquid hot water processes generally liberate the sugars in an oligomeric form and thus require a secondary acidic or enzymatic hydrolysis step to produce fermentable monomeric sugars.

Another approach for liquid hot water pretreatment uses some chemicals as agents to control the pH in the range of 4–7 (10, 46). With some feedstocks, such as corn stover, there may be enough inherent buffering capacity from the feedstock that the target pH range is achieved without any requirement of pH-controlling chemicals. The reason for controlling the pH is to retain the released hemicellulose sugars in oligomeric form as a means of controlling sugar degradation losses and fermentation inhibitor formation (10). However, recent data suggests that this approach may not be effective in achieving high enzymatic digestibility of the cellulose (>80% glucose yield from the available cellulose) in pretreated corn stover (17). In general, liquid hot water pretreatments are attractive from a process cost-savings potential (no catalyst usage, low-cost reactor construction due to low-corrosion potential), but these cost savings can be offset by lower overall sugar yields and the need for enzymatic hydrolysis conversion steps (21).

14.5.3.2 Liquid hot water percolation pretreatment

Pressurized liquid hot water that is percolated or otherwise forced through a packed bed of biomass particles has been shown to result in high removal of both hemicellulose and lignin, with high recovery of hemicellulose-derived sugars (primarily in oligomeric form) and high digestibility of the resulting pretreated solids (47, 48). These processes may be difficult to commercialize due to the high volumes of liquid required to sustain a continuously-flowing percolation process, although some efforts to address the high liquid volume requirements using intermittent-flow approaches have been investigated (16). Nevertheless, percolation pretreatment techniques are useful in research applications to generate pretreated solids with a wide range of hemicellulose and lignin removal extents for enzymatic hydrolysis and related compositional and ultrastructure studies.

14.5.4 Acidic pretreatments

14.5.4.1 Dilute acid batch/co-current pretreatment

Dilute acid pretreatment is probably the most thoroughly investigated biomass pretreatment technique. A variety of acidic catalysts have been investigated in numerous batch/co-current dilute acid pretreatment reactor designs on a wide range of woody, herbaceous, and agricultural residue feedstocks. For cost reasons, most dilute acid pretreatment studies have utilized sulfuric acid or gaseous sulfur dioxide (in steam explosion applications), although several processes that utilize nitric, phosphoric, hydrochloric, or carbonic acid have also been investigated. Dilute acid pretreatment studies find wide distribution in the published literature and have been summarized in several review articles (10–15, 49, 50).

Dilute acid batch and co-current pretreatments are generally aimed at achieving near-complete solubilization of the hemicellulose fraction of biomass while also achieving high yields of hemicellulose-derived sugars. Many processes seek to directly achieve monomeric sugar formation, although care must be taken to prevent excessive sugar degradation of monomeric sugars. If performed properly, dilute acid pretreatment can be very effective at achieving reasonable monomer sugar yields via both hemicellulose hydrolysis and enzymatic digestion of the cellulose in the resulting solids across a range of biomass feedstock types (51). However, in batch or co-current mode, there will likely be some degradation losses of hemicellulose-derived sugars and possibly a requirement for conditioning the hydrolyzate liquid fraction prior to fermentation. Dilute acid pretreatment approaches have been tested in continuous co-current pilot scale reactor systems (52, 53) and have been the subject of intensive process simulation and economic analysis for potential commercial-scale operations using this pretreatment approach (54).

14.5.4.2 Dilute acid percolation/countercurrent pretreatment

For reasons similar to liquid hot water percolation processes, dilute acid processes that employ a percolation mode of operation have also been investigated. Very high yields of monomeric and oligomeric xylose have been obtained in a two-stage percolation pretreatment of hardwoods, with high enzymatic hydrolysis yields of the cellulose in the pretreated solids (55). The high digestibility achieved in this approach has been attributed to significant lignin solubilization and removal from the pretreated solids in the continuously-flowing percolation process.

Kinetic modeling studies and associated experimental work have shown that extension of the percolation concept to a countercurrent contacting of biomass particles with the flowing dilute acid medium can further reduce hemicellulose-derived sugar degradation losses and also produce highly digestible pretreated solids. This is achieved by further reduction of the residence time of released sugars under reaction conditions based on the observed first-order hydrolysis reaction kinetics (56). This concept has also been extended to a full thermochemical hydrolysis of both hemicellulose and cellulose, with much higher sugar yields than traditional batch or co-current two-stage dilute acid hydrolysis processes (56). Such processes are highly attractive from a sugar yield standpoint, but will be difficult to apply commercially due to the high liquid volume requirements and complex large-scale reactor configurations.

Dilute acid percolation and countercurrent processes that use lower liquid volumes yet still achieve the highly digestible pretreated solids attributed to lignin solubilization, have also been investigated. In this approach, pretreatment is conducted in a batch mode, followed by a separation and a limited-volume washing of the pretreated solids prior to cooling below the lignin phase-transition temperature, where re-precipitation of solubilized lignin would be expected to occur. High enzymatic digestibility has been reported using this approach on a yellow poplar hardwood feedstock (57), but further work revealed limited benefit to this approach using corn stover as a feedstock.

14.5.5 Alkaline pretreatments

14.5.5.1 Sodium hydroxide pretreatment

Alkali pretreatment processes generally do not hydrolyze hemicellulose as extensively as acidic pretreatments, but can be effective at removing lignin, which can lead to an increase in the enzymatic digestibility of alkali pretreated solids. Several studies on alkali pretreatment using sodium hydroxide have been reported and reviewed (11, 12, 14). This pretreatment approach causes swelling of fibers, leading to an increase in internal surface area, reduction in the degree of polymerization, a decrease in crystallinity, separation of the structural linkages between lignin and carbohydrates, and disruption of lignin structure (11). The effectiveness of sodium hydroxide pretreatment has been correlated to feedstock lignin content, with high lignin feedstocks, especially softwoods, showing poor performance using this approach (14). Dilute sodium hydroxide pretreatment has been shown to be quite effective on low lignin (10–18% lignin content) straw feedstocks (58), but the cost effectiveness of this pretreatment approach has not been thoroughly evaluated.

14.5.5.2 Ammonia pretreatment

In addition to the rapid expansion AFEX pretreatment process, which utilizes ammonia to achieve both chemical and physical changes to biomass, there are a number of additional ammonia pretreatment processes. The simplest ammonia pretreatment process involves a relatively low-temperature soaking (ambient temperature up to 90°C) using aqueous ammonia (various strengths up to 29 wt% NH_4OH) at solids loadings of 10–50% and residence times from a few hours to up to 1 day (14, 59, 60). In these processes, up to 80% delignification has been reported on feedstocks such as wheat straw and corn stover, with much lower extents of hemicellulose solubilization. However, good enzymatic digestibility of the remaining cellulose and some of the remaining hemicellulose can be achieved using commercial cellulase preparations (60). There is little evidence that this pretreatment approach would be effective on woody biomass types, especially softwoods. As with other alkaline pretreatment approaches, the augmentation of cellulase enzymes with hemicellulase and other accessory enzyme activities could improve the enzymatic saccharification of ammonia-pretreated biomass.

A percolation-type ammonia pretreatment process, known as ammonia-recycled percolation (ARP), has also been investigated (10, 14, 19, 61). The ARP process passes dilute aqueous ammonia (<15% NH_3) through a packed bed of biomass at temperatures of 150–170°C. Because of the flow-through nature and ammonia-based chemistry, ARP pretreatment of

corn stover can achieve very high lignin removal (above 80% delignification) and moderate hemicellulose solubilization (above 50% xylan solubilization at residence times of 20 min or more), although hemicellulosic sugars are generally recovered in oligomeric form and would require additional processing to liberate monomeric sugars. Enzymatic digestibility of ARP-pretreated corn stover is also high (about 90% conversion of residual cellulose to glucose and about 70% conversion of residual xylan to xylose) using a commercial cellulase preparation (19). However, economic analysis has revealed that the high liquid volumes and subsequent dilute process streams does not allow the ARP process to economically compete with other pretreatments, even when efficient ammonia recovery and recycle is assumed (21). In this respect, ARP is similar to liquid hot water and dilute acid percolation pretreatments in that such processes may not be economically competitive, but they are of value in research applications to generate pretreated solids with a wide range of hemicellulose and lignin removal extents for enzymatic hydrolysis and related compositional and ultrastructure studies.

14.5.5.3 Lime pretreatment

Pretreatment using lime has been studied as a low-cost process that primarily achieves acetyl and lignin solubilization (10, 14, 18, 62, 63). Lime pretreatment has been practiced at a wide range of temperatures, from 25°C to about 130°C, with lime loadings of about 10 wt% (on a dry feedstock basis) and solids loadings of 20% or less. At the higher temperatures, the pretreatment times are reasonably short (minutes to hours), but can extend to several weeks at lower temperatures. Because of the lengthy residence time at low temperatures, lime pretreatment can be conducted in a pile arrangement without expensive pressure reactors and can be performed as part of the feedstock storage system (18). Near-complete deacetylation generally occurs upon lime pretreatment of low-lignin herbaceous feedstocks and agricultural residues, with about 30% lignin removal. Much higher lignin removal (up to 80%) can be achieved by adding oxygen or air to the lime pretreatment system. In the lower temperature lime pretreatment pile arrangement, this can be accomplished by percolating air through the pretreatment pile. The additional lignin removal under oxidative conditions allows lime pretreatment to achieve reasonable enzymatic digestibility using more recalcitrant feedstocks, such as sugar cane bagasse and hardwoods (14). Lime pretreatment generally requires longer pretreatment times than other alkaline pretreatments, such as those using ammonia, but catalyst costs are lower. However, regeneration and reuse of the lime will probably still be necessary, and such recovery systems will add significant capital and operating costs to the lime pretreatment approach (21).

14.5.6 Solvent pretreatments

14.5.6.1 Organic solvents

Numerous organic or organic-aqueous solvent mixtures utilizing methanol, ethanol, acetone, ethylene glycol, triethylene glycol, and tetrahydrofurfuryl alcohol have been used as biomass pretreatment processes to solubilize lignin (10–13, 64, 65). Such processes are commonly referred to as organosolv processes. In some studies, inorganic acid catalysts, such as sulfuric or hydrochloric acid, are added to achieve significant levels of hemicellulose hydrolysis and even cellulose hydrolysis (66) along with lignin solubilization. In some cases,

the main components of biomass (cellulose, hemicellulose, and lignin) can be effectively fractionated, with each component potentially used for separate value-added products (67). Solvents must be effectively recovered and recycled using appropriate extraction and separation techniques without leaving behind any inhibitory levels of residual solvents in process streams that undergo subsequent biological processing. While residual cellulose-rich pretreated solids from such processes may be highly digestible using cellulase enzymes, the cost of such processes and the potential value of the relatively pure fractions may make them better suited to higher-value applications.

14.5.6.2 Cellulose-dissolving solvents

Another category of solvent pretreatment involves the use of cellulose-dissolving solvents, such as cadoxen, concentrated mineral acids, DMSO, and zinc chloride (10, 12). While these agents can be effective at directly releasing sugars from the carbohydrate fractions of biomass and/or producing a solid residue containing cellulose that is highly digestible by enzymes, the use of such solvents in pretreatment processes for the production of fuels and commodity chemicals from biomass will be challenging due to the expense of such catalysts, catalyst recycle requirements, and the requirement for clean process streams for subsequent biological conversions.

14.5.7 Supercritical fluid pretreatments

Biomass pretreatment processes using supercritical fluids to extract lignin from biomass feedstock have been investigated. A number of different supercritical fluids (alone or in mixtures) have been investigated, although the most common approaches utilize water, carbon dioxide, or ammonia (14, 68). While supercritical pretreatment conditions can effectively remove lignin and produce pretreated biomass that exhibits good enzymatic digestibility, the economic viability and practical operation of processes at supercritical operating conditions have not been effectively demonstrated. Of greatest concern are the extremely high-pressure requirements (generally above 10 MPa) of these processes.

14.5.8 Oxidative pretreatments

Oxidative processes for biomass pretreatment applications are often referred to as wet oxidation processes. This approach was born out of efforts in the pulp and paper industry to develop oxygen delignification processes to reduce chlorine use in pulping. The most common approach for wet oxidation as a biomass pretreatment involves the injection of pressurized O₂ into a pretreatment reactor at temperatures up to 200°C and pressures up to about 1.5 MPa (69). Another approach utilizes a percolation-type pretreatment that incorporates wet oxidation, among other approaches, into a biomass fractionation process (70). Most work has been performed in batch reactors at low solids loading, occasionally with the use of small amounts of co-catalysts or solvents. Much of this work has included the use of alkaline buffers (usually sodium carbonate) to maintain reaction pH in the neutral to alkaline range. There have been sporadic efforts to investigate alkaline peroxides and ozone as other types of oxidizing agents (71–73).

Wet oxidation extensively delignifies biomass with production of monomeric and oligomeric phenols, followed by oxidative cleavage to a variety of carboxylic acids. When the reaction is not buffered and pH is allowed to drift naturally down, extensive formation of furfurals occurs, which can also be cleaved to form carboxylic acids in the oxidative environment. Hemicellulose is typically solubilized to about 70% conversion, primarily as oligomers. The combination of extensive delignification and at least 50% hemicellulose removal can result in highly digestible pretreated solids (70).

14.5.9 Biological pretreatment

Most biological pretreatment approaches utilize certain classes of lignin-solubilizing microorganisms that will produce a lignocellulosic feedstock that is more amenable to enzymatic saccharification than native biomass. Many studies have focused on a class of microorganisms known as white-rot fungi, which produce lignin-degrading enzymes, lignin peroxidases, peroxide producing enzymes, polyphenol oxidases, laccases, and quinone-reducing enzymes (11, 12, 74, 75). In such biological pretreatment processes, biomass is inoculated with appropriate fungal cultures and incubated for several weeks, followed by evaluation of the enzymatic hydrolyzability of the treated biomass. While increased digestibility has been attributed to the delignification action of these microorganisms (75, 76), there is often some parallel loss of cellulose and/or hemicellulose during the biological pretreatment.

Other biological pretreatment approaches have been investigated as a means of preparing stored biomass for subsequent mechanical, thermal or chemical pretreatment, potentially reducing the required severity and cost of the pretreatment process. This approach can effectively soften especially woody biomass types, reducing the mechanical power requirement for size reduction via disk refining by about 40%, although there is some associated loss of mass due to fungal action of lignin and structural carbohydrates, suggesting a potential loss of available carbohydrates to the conversion process (77).

14.6 Future prospects

To further decrease the cost of the pretreatment part of any biomass conversion process it is essential that sugar yields are increased (sugar losses are minimized), solids concentrations are as high as possible, and the capital cost of the pretreatment reactors and associated equipment are kept as low as possible. It is also desirable that the pretreatment process increases the ease with which the cellulose in the pretreated biomass can be saccharified so that it can be accomplished with lower cellulase loadings or in a shorter digestion time.

Development and improvement of existing pretreatment processes is the subject of intensive research at several institutions in the USA and overseas. One approach that is receiving more attention is the investigation of the effects of pretreatment at a more fundamental level; increasingly research is looking at the effects of biomass pretreatment at the cellular, ultrastructural and even molecular level of the plant cell wall. The plant cell wall is highly complex at all length scales and especially chemically heterogeneous at the molecular level. For example, in nature dozens of glycosyl hydrolases are involved in plant cell wall deconstruction. These glycosyl hydrolases act on cell wall polysaccharides specifically and synergistically; however, the biochemistry of these unique catalytic events, caused by

individual enzymes acting on biomass as well as the consequences of thermal chemical pretreatment on these reactions, remains poorly understood. There is an increasing focus on characterization of biomass's molecular structure and ultrastructure in order to gain sufficient understanding of the relationships between pretreatment chemistry and enzyme digestibility. There is also a desire to determine the effects that cellulases and other enzymes have on biomass ultrastructure.

Researchers at Michigan State University have been studying the changes that occur at the cellular level in corn stover upon AFEX pretreatment (78). After staining with a lignin-specific dye, safranin, they observed a change in the distribution of lignin in the cells by confocal laser microscopy. Scanning electron micrographs also appeared to indicate a change in distribution of lignin-like compounds and they postulated that the ammonia partially cleaved the lignin resulting in a decrease in its glass transition temperature so that the lignin could be relatively easily mobilized at temperatures close to 100°C. They also saw a 10–30% decrease in the oxygen:carbon ratio by X-ray photoelectron spectroscopy (XPS) at the surface of the cells indicating an increase in carbon-rich species such as lignin. A redistribution of lignin to the surface of plant cell walls caused by pretreatment could have a significant effect on the digestibility of the cellulose.

Similar effects have been observed in dilute acid pretreatment of corn stover. Dilute acid pretreatments are performed at much higher temperatures than AFEX pretreatments (140–200°C versus about 90°C). The level of lignin redistribution is therefore much higher. In dilute acid pretreatment, droplets are seen to form that appear to be at least lignin derived. The droplets appear in the liquid hydrolyzate, adhere to the surface of the biomass and also collect in spaces between cells (Figure 14.1). Visually, these microscopic changes appear very significant; however, it is not known at this time how significant they are in altering the digestibility or accessibility of the cellulose in the pretreated materials.

The plant cell wall microfibril, the primary target substrate of bioconversion, is believed to consist of a cellulose-elementary-fibril core surrounded by a hemicellulose sheath forming a macromolecular composite (79). In acidic pretreatments hydrolysis and solubilization of the

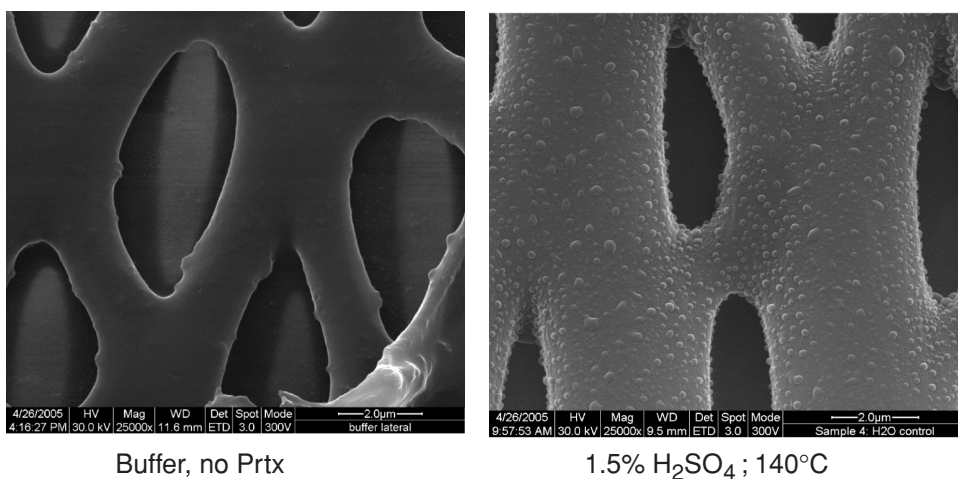


Figure 14.1 SEM imaging of xylem vessel interior (pores) showing putative lignin containing droplets formed during dilute acid hydrolysis.

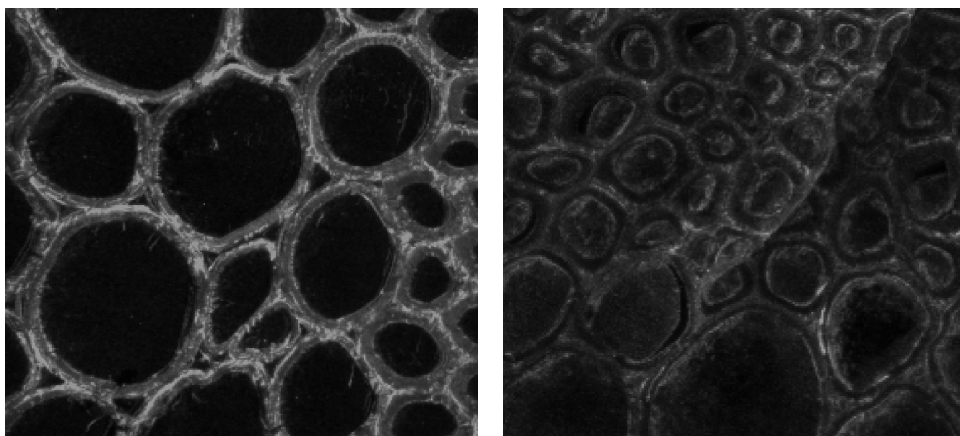


Figure 14.2 CLSM micrographs of unpretreated (left) and dilute acid pretreated (160°C, 10 min, 87% xylan removal) corn stover rind (right) sections labeled with LM11 α -xylan antibody that is bound to a secondary antibody conjugated to a fluorescent protein (Alexa488) showing decrease in fluorescent signal due to xylan hydrolysis (imaging by Stephanie Porter). (Reproduced in color as Plate 29).

hemicellulose is thought to be the main mechanism by which accessibility of the cellulose to enzymes is increased. Changes in xylan distribution are being studied by probing pretreated materials with xylan-specific monoclonal antibodies. The distribution of xylan in cells and within cell walls can then be imaged by binding the antibody with a secondary fluorescent dye (Alexa-488) for confocal laser scanning microscopy (CLSM) or a gold nanoparticle for transmission electron microscopy (TEM). In dilute acid pretreatment of corn stover rind differences in the pattern of fluorescence have been observed that are attributable to changes in the distribution of xylan in the cell wall (Figure 14.2). The cause of this is still being investigated and the effect of differences in xylan distribution on cellulose digestibility is still being interpreted.

It is anticipated that studies such as those described above will lead to significant advances in our understanding of how pretreatment can overcome the natural recalcitrance of biomass and increase the digestibility of biomass in cost-effective pretreatments. An improved understanding should then allow for design of improved pretreatments that decrease the cost of converting biomass feedstocks into fermentable monomeric sugars.

Acknowledgment

This work was supported by the US DOE Office of the Biomass Program.

References

1. Springfield, R.M. & Hester, R.D. (2001) Development and modeling of a continuous simulated moving bed ion exclusion process for the separation of acid and sugar. *Separation Science and Technology*, **36**, 911.
2. Saha, B.C. (2003) Hemicellulose bioconversion. *Journal of Industrial Microbiology and Biotechnology*, **30**, 279.

3. Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y., Holtzapple, M. & Ladisch, M. (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology*, **96**, 673.
4. Coughlan, M.P. (1991) Enzymic hydrolysis of cellulose: An overview. *Bioresource Technology*, **39**, 107.
5. Mansfield, S.D., Mooney, C. & Saddler, J.N. (1999) Substrate and enzyme characteristics that limit cellulose hydrolysis. *Biotechnology Progress*, **15**, 804.
6. Esteghlalian, A.R., Srivastava, V., Gilkes, N.R., Gregg, D.J. & Saddler, J.N. (2001) An overview of factors influencing the enzymatic hydrolysis of lignocellulosic feedstocks. In: *ACS Symposium Series 769, Glycosyl Hydrolases in Biomass Conversion* (eds M.E. Himmel, J.O. Baker & J.N. Saddler). American Chemical Society, Washington, DC.
7. Converse, A.O. (1993) Substrate factors limiting enzymatic hydrolysis. In: *Bioconversion of Forest and Agricultural Plant Residues* (ed. J.N. Saddler). CAB International, Oxford, UK.
8. Jeoh, T., Ishizawa, C.I., Davis, M.F., Himmel, M.E., Adney, W.S. & Johnson, D.K. (2007) Cellulase digestibility of pretreated biomass is limited by cellulose accessibility. *Biotechnology and Bioengineering*, **98**, 112.
9. Zhang, Y.H.P. & Lynd, L.R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering*, **88**, 797.
10. Sun, Y. & Cheng, J. (2002) Hydrolysis of lignocellulosic materials for ethanol production: A review. *Bioresource Technology*, **83**, 1.
11. Hsu, T.A. (1996) Pretreatment of biomass. In: *Handbook on Bioethanol, Production and Utilization* (ed. C.E. Wyman). Taylor & Francis, Washington, DC.
12. Duff, S.J.B. & Murray, W.D. (1996) Bioconversion of forest products industry waste cellulose to fuel ethanol: A review. *Bioresource Technology*, **55**, 1.
13. McMillan, J.D. (1994) Pretreatment of lignocellulosic biomass. In: *Enzymatic Conversion of Biomass for Fuels Production* (eds M.E. Himmel, J.O. Baker & R.P. Overend). American Chemical Society, Washington, DC.
14. Wyman, C.E., Dale, B.E., Elander, R.T., Holtzapple, M., Ladisch, M.R. & Lee, Y.Y. (2005) Coordinated development of leading biomass pretreatment technologies. *Bioresource Technology*, **96**, 1959.
15. Lloyd, T.A. & Wyman, C.E. (2005) Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids. *Bioresource Technology*, **96**, 1967.
16. Liu, C. & Wyman, C.E. (2005) Partial flow of compressed-hot water through corn stover to enhance hemicellulose sugar recovery and enzymatic digestibility of cellulose. *Bioresource Technology*, **96**, 1978.
17. Mosier, N., Hendrickson, R., Ho, N., Sedlak, M. & Ladisch, M.R. (2005) Optimization of pH controlled liquid hot water pretreatment of corn stover. *Bioresource Technology*, **96**, 1986.
18. Kim, S. & Holtzapple, M.T. (2005) Lime pretreatment and enzymatic hydrolysis of corn stover. *Bioresource Technology*, **96**, 1994.
19. Kim, T.H. & Lee, Y.Y. (2005) Pretreatment and fractionation of corn stover by ammonia recycle percolation process. *Bioresource Technology*, **96**, 2007.
20. Teymouri, F., Laureano-Perez, L., Alizadeh, H. & Dale, B.E. (2005) Optimization of ammonia fiber explosion (AFEX) treatment parameters for enzymatic hydrolysis of corn stover. *Bioresource Technology*, **96**, 2014.
21. Eggeman, T. & Elander, R.T. (2005) Process and economic analysis of pretreatment technologies. *Bioresource Technology*, **96**, 2019.
22. Wyman, C.E., Dale, B.E., Elander, R.T., Holtzapple, M., Ladisch, M.R. & Lee, Y.Y. (2005) Comparative sugar recovery data from laboratory scale application of leading pretreatment technologies to corn stover. *Bioresource Technology*, **96**, 2026.

23. Millett, M.A., Effland, M.J. & Caulfield, D.L. (1979) Influence of fine grinding on the hydrolysis of cellulosic materials – acid versus enzymes. *Advances in Chemistry Series*, **181**, 71.
24. Fan, L.T., Lee, Y. & Gharpuray, M.M. (1982) The nature of lignocellulosics and their pretreatments for enzymatic hydrolysis. *Advances in Biochemical Engineering*, **23**, 157.
25. Tassinari, T., Macy, C. & Spano, L. (1980) Energy requirements and process design considerations in compression-milling pretreatment of cellulosic wastes for enzymatic hydrolysis. *Biotechnology and Bioengineering*, **22**, 1689.
26. Tassinari, T., Macy, C. & Spano, L. (1982) Technology advances for continuous compression milling pretreatment of lignocellulosics for enzymatic hydrolysis. *Biotechnology and Bioengineering*, **24**, 1495.
27. Cadoche, L. & Lopez, G.D. (1989) Assessment of size reduction as a preliminary step in the production of ethanol from lignocellulosic wastes. *Biological Wastes*, **30**, 153.
28. Ooshima, H., Aso, K., Harano, Y. & Yamamoto, T. (1984) Microwave treatment of cellulosic materials for their enzymatic hydrolysis. *Biotechnology Letters*, **6**, 289.
29. Mason, W.H. (1926) Process and apparatus for disintegration of wood and the like. U.S. Patent 1,578,609.
30. Saddler, J.N., Ramos, L.P. & Breuil, C. (1993) Steam pretreatment of lignocellulosic residues. In: *Bioconversion of Forest and Agricultural Plant Wastes* (ed. J.N. Saddler). CAB International, Wallingford, UK.
31. Brownell, H.H. & Saddler, J.N. (1987) Steam pretreatment of lignocellulosic material for enhanced enzymatic-hydrolysis. *Biotechnology and Bioengineering*, **29**, 228.
32. Wright, J.D. (1988) Ethanol from biomass by enzymatic hydrolysis. *Chemical Engineering Progress*, **84**, 62.
33. Mackie, K.L., Brownell, H.H., West, K.L. & Saddler, J.N. (1985) Effect of sulphur dioxide and sulphuric acid on steam explosion of aspenwood. *Journal of Wood Chemistry and Technology*, **5**, 405.
34. Saddler, J.N., Brownell, H.H., Clermont, L.P. & Levitin, N. (1982) Enzymatic hydrolysis of cellulose and various pretreated wood fractions. *Biotechnology and Bioengineering*, **24**, 1389.
35. Clark, T.A. & Mackie, K.L. (1987) Steam explosion of the softwood *Pinus radiata* with sulphur dioxide addition. I. Process optimization. *Journal of Wood Chemistry and Technology*, **7**, 373.
36. Heitz, M., Capek-Menard, E., Koeberle, P.G., Gangne, J. & Chornet, E. (1991) Fractionation of *Populus tremuloides* at the pilot plant scale: Optimization of steam pretreatment conditions using the Stake II technology. *Bioresource Technology*, **35**, 23.
37. Jollez, P., Chornet, E. & Overend, R.P. (1994) Steam-aqueous fractionation of sugar cane bagasse: An optimization study of process conditions at the pilot plant level. *Advances in Thermochemical Biomass Conversion*, **2**, 1659.
38. Dale, B.E. & Moreira, M.J. (1982) A freeze explosion technique for increasing cellulose hydrolysis. *Biotechnology and Bioengineering Symposium*, **12**, 31.
39. Dale, B.E., Leong, C.K., Pham, T.K., Esquivel, V.M., Rios, I. & Latimer, V.M. (1996) Hydrolysis of lignocellulosics at low enzyme levels: Application of the AFEX process. *Bioresource Technology*, **56**, 111.
40. Moniruzzaman, M., Dale, B.E., Hespell, R.B. & Bothast, R.J. (1997) Enzymatic hydrolysis of high moisture content corn fiber pretreated by AFEX and recovery and recycle of the enzyme complex. *Applied Biochemistry and Biotechnology*, **67**, 113.
41. Gollapalli, L.E., Dale, B.E. & Rivers, D.M. (2002) Predicting digestibility of ammonia fiber explosion (AFEX)-treated rice straw. *Applied Biochemistry and Biotechnology*, **98–100**, 23.
42. Dale, B.E., Weaver, J. & Byers, F.M. (1999) Extrusion processing for ammonia fiber explosion (AFEX). *Applied Biochemistry and Biotechnology*, **77–79**, 1.
43. Bobleter, O. (1994) Hydrothermal degradation of polymers derived from plants. *Progress in Polymer Science*, **20**, 2083.

44. Mok, W.S.L. & Antal, M.J., Jr. (1992) Uncatalyzed solvolysis of whole biomass hemicellulose by hot compressed liquid water. *Industrial and Engineering Chemistry Research*, **31**, 1157.
45. Allen, S.G., Schulman, D., Lichwa, J., Antal, M.J., Jr., Jennings, E. & Elander, R. (2001) A comparison of aqueous and dilute-acid single-temperature pretreatment of yellow poplar sawdust. *Industrial and Engineering Chemistry Research*, **40**, 2352.
46. Weil, J.R., Brewer, M., Hendrickson, R., Sarikaya, A. & Ladsich, M.R. (1997) Continuous pH monitoring during pretreatment of yellow poplar sawdust by pressure cooking in water. *Applied Chemistry and Biotechnology*, **68**, 21.
47. Allen, S.G., Kam, L.C., Zeman, A.J. & Antal, M.J., Jr. (1996) Fractionation of sugar cane with hot compressed liquid water. *Industrial and Engineering Chemistry Research*, **35**, 2709.
48. Liu, C. & Wyman, C.E. (2003) The effect of flow rate of compressed hot water on xylan, lignin, and total mass removed from corn stover. *Industrial and Engineering Chemistry Research*, **42**, 5409.
49. Jacobsen, S.E. & Wyman, C.E. (1999) Hemicellulose and cellulose hydrolysis models for application to current and novel pretreatment processes. *Applied Biochemistry and Biotechnology*, **84–86**, 81.
50. Lee, Y.Y., Iyer, P. & Torget, R.W. (1999) Dilute-acid hydrolysis of lignocellulosic biomass. *Advances in Biochemical Engineering and Biotechnology*, **65**, 93.
51. Esteghalian, A., Hashimoto, A.G., Fenske, J.J. & Penner, M.H. (1997) Modeling and optimization of dilute-sulfuric-acid pretreatment of corn stover, poplar, and switchgrass. *Bioresource Technology*, **59**, 129.
52. Tucker, M.P., Farmer, J.D., Keller, F.A., Schell, D.J. & Nguyen, Q.A. (1998) Comparison of yellow poplar pretreatment between NREL digester and Sunds hydrolyzer. *Applied Biochemistry and Biotechnology*, **70–72**, 25.
53. Schell, D.J., Farmer, J., Newman, M. & McMillan, J.D. (2003) Dilute-sulfuric acid pretreatment of corn stover in pilot-scale reactor – Investigation of yields, kinetics, and enzymatic digestibilities of solids. *Applied Biochemistry and Biotechnology*, **105**, 69.
54. Aden, A., Ruth, M., Ibsen, K., Jechura, J., Neeves, K., Sheehan, J., Wallace, B., Montague, L., Slayton, A. & Lukas, J. (2002) *Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Acid Prehydrolysis for Corn Stover*. National Renewable Energy NREL/TP-510-32438, Golden, CO.
55. Torget, R., Hatzis, C., Hayward, T.K., Hsu, T.A. & Philippidis, G.P. (1996) Optimization of reverse-flow, two-temperature, dilute-acid pretreatment to enhance biomass conversion to ethanol. *Applied Biochemistry and Biotechnology*, **57–58**, 85.
56. Lee, Y.Y., Wu, Z.W. & Torget, R.W. (2000) Modeling of countercurrent shrinking-bed reactor in dilute-acid total-hydrolysis of lignocellulosic biomass. *Bioresource Technology*, **71**, 29.
57. Nagle, N.J., Elander, R.T., Newman, M.N., Rohrback, B.T., Ruiz, R.O. & Torget, R.W. (2002) Efficacy of a hot washing process for pretreated yellow poplar to enhance bioethanol production. *Biotechnology Progress*, **18**, 734.
58. Bjerre, A.B., Olesen, A.B. & Fernqvist, T. (1996) Pretreatment of wheat straw using combined wet oxidation and alkaline hydrolysis resulting in convertible cellulose and hemicellulose. *Biotechnology and Bioengineering*, **49**, 568.
59. Detroy, R.W., Lindenfelser, L.A., St Julian, J.G. & Orton, W.L. (1980) Saccharification of wheat-straw cellulose by enzymatic hydrolysis following fermentative and chemical pretreatment. *Biotechnology and Bioengineering Symposium Series*, **10**, 135.
60. Kim, T.H. & Lee, Y.Y. (2005) Pretreatment of corn stover by soaking in aqueous ammonia. *Applied Biochemistry and Biotechnology*, **121**, 1119.
61. Iyer, P.V., Wu, Z., Kim, S.B. & Lee, Y.Y. (1996) Ammonia recycled percolation process for pretreatment of herbaceous biomass. *Applied Biochemistry and Biotechnology*, **57–58**, 121.
62. Kaar, W.E. & Holtzaple, M.T. (2000) Using lime pretreatment to facilitate the enzymatic hydrolysis of corn stover. *Biomass and Bioenergy*, **18**, 189.

63. Chang, V.S., Nagwani, M., Kim, C.H. & Holtzapple, M.T. (2001) Oxidative lime pretreatment of high-lignin biomass. *Applied Biochemistry and Biotechnology*, **94**, 1.
64. Chum, H.L., Johnson, D.K., Black, S., Baler, J., Grohmann, K. & Sarkanen, K.V. (1988) Organosolv pretreatment for enzymatic hydrolysis of poplars: I. Enzyme hydrolysis of cellulosic residues. *Biotechnology and Bioengineering*, **31**, 643.
65. Thring, R.W., Chornet, E. & Overend, R. (1990) Recovery of a solvolytic lignin: Effects of spent liquor/acid ratio, acid concentration and temperature. *Biomass*, **23**, 289.
66. Paszner, L. & Cho, H.J. (1988) High efficiency of lignocellulosics to sugars for liquid fuel production by the ACOS process. *Energy Exploitation Exploration*, **6**, 39.
67. Black, S.K., Hames, B.R. & Myers, M.D. (1998) Method of separating lignocellulosic material into lignin, cellulose, and dissolved sugars. U.S. Patent 5,730,837.
68. Li, L. & Koran, E. (1988) Interaction of supercritical fluids with lignocellulosic materials. *Industrial and Engineering Chemistry Research*, **27**, 1301.
69. Thomsen, A. & Schmidt, A.S. (1999) *Further Development of Chemical and Biological Processes for Production of Bioethanol: Optimization of Pretreatment Processes and Characterization of Products*. Riso National Laboratory Report R-1110(EN), Roskilde, Denmark.
70. Wingerson, R. (2002) High efficiency organosolv saccharification process. U.S. Patent 6,419,788.
71. Saha, B.C. & Cotta, M.A. (2006) Ethanol production from alkaline peroxide pretreated enzymatically saccharified wheat straw. *Biotechnology Progress*, **22**, 449.
72. Yang, B., Boussaid, A., Mansfield, S., Gregg, D. & Saddler, J.N. (2002) Fast and efficient alkaline peroxide treatment to enhance the enzymatic digestibility of steam-exploded softwood substrates. *Biotechnology and Bioengineering*, **77**, 678.
73. Silverstein, R.A., Chen, Y., Sharma-Shivappa, R.R., Boyette, M.D. & Osborne, J. (2007) A comparison of chemical pretreatment methods for improving saccharification of cotton stalks. *Bioresource Technology*, **98**, 3000.
74. Blanchette, R.A. (1991) Delignification by wood-decay fungi. *Annual Reviews in Phytopathology*, **29**, 381.
75. Akin, D.E., Rigsby, L.L., Sethuraman, A., Morrison, W.H., III, Gamble, G.R. & Eriksson, K.E.L. (1995) Alterations in structure, chemistry, and biodegradability of grass lignocellulose treated with the white rot fungi *Ceriporiopsis subvermispora* and *Cyanthus stercoreus*. *Applied and Environmental Microbiology*, **61**, 1591.
76. Hatakka, A.I., Keller, F.A., Hamilton, J.E. & Nguyen, Q.A. (1983) Pretreatment of wheat straw by white-rot fungi for enzymatic saccharification of cellulose. *Applied Microbiology and Biotechnology*, **18**, 350.
77. Keller, F.A., Hamilton, J.E. & Nguyen, Q.A. (2003) Microbial pretreatment of biomass: Potential for reducing severity of thermochemical biomass pretreatment. *Applied Biochemistry and Biotechnology*, **105–108**, 27.
78. Balan, V., Chundawat, S., Bals, B. & Dale, B. (2006) Fundamental understanding of biomass pretreatment technologies: The case of Ammonia Fiber Expansion (AFEX), November 17, 2006. Available from: <http://www.everythingbiomass.org>.
79. Ding, S.Y. & Himmel, M.E. (2006) The maize primary cell wall microfibril: A new model derived from direct visualization. *Journal of Agricultural and Food Chemistry*, **54**, 597.

Chapter 15

Understanding the Biomass Decay Community

William S. Adney, Daniel van der Lelie, Alison M. Berry, and Michael E. Himmel

15.1 Introduction

Traditionally, microbiologists have taken a reductionist approach to understanding microbial biomass decay communities focusing on the analysis of individual genes, microorganisms, and biochemical reactions. Recent advances in molecular biology have identified many genetic components but have provided limited information on the mechanisms of biomass decay. In the area of biochemistry, advances in proteomics and high-throughput enzyme assays are providing new theories into the mechanisms of biomass decay, but limited information on individual microorganisms and their interactions. Combined, however, these technologies are providing a new “systems biology” approach to understanding the biomass decay communities. This approach will allow microbiologists to envision and model microbial biomass decay as a set of interacting processes that when combined effectively degrade plant biomass.

The first major step towards clarifying the fundamental principles of biomass recalcitrance is to understand the scale and complexity of natural systems involved in biomass decomposition. Heterotrophic microorganisms are major players in biomass decomposition and in the global cycling of terrestrial carbon. Our terrestrial biosphere depends on heterotrophs functioning in complex and dynamic communities to breakdown the natural accumulation of biomass. Although the subject of study for several decades, we still know little about the diversity and complex interrelationships of the individual organisms. However, we now understand that these communities vary in spatial and temporal dimensions to control the biochemical rate of carbon cycling, as well as the cycling of other essential elements, like nitrogen, sulfur, and phosphorus. In fact, the production of carbon dioxide by chemoorganotrophs is the single most important contribution of CO₂ to the atmosphere (1).

Microbial communities are complex networks of individual organisms that include every ecological relationship ever described, ranging from coexistence to commensalism, mutualism, and parasitism. There are direct symbioses between individual microorganisms and indirect symbioses in which metabolic processes of one species modify the habitat and/or physiology of another species. Studying microbial communities in most environments has

thus presented a challenge in that it is very difficult to simulate environmental conditions and ecological relationships adequately in the laboratory in order to satisfy the physiological requirements for the reproducible cultivation of a representative community. This may be due to our inability to reproduce the physical, chemical, and temporal conditions needed for the multiplex interactions and unknown species.

Until recently, biodiversity estimates were based only on those species that could be cultivated by using traditional *in vitro* microbiology techniques. Unfortunately, these techniques only allow 1–5% of the total community members to be examined (2, 3). Traditional microbiology techniques were developed to study the growth and metabolic requirements of individual organisms in pure culture. As such, they provide limited information into the actions of biomass-degrading communities determined by diverse and dynamic biochemical pathways. Recently, new molecular technologies have provided valuable information about the *in situ* biodiversity of plant decaying microbial communities. However, while genomic approaches provide important information about the diversity of individual species within populations they do not predict the biochemical outcome on functional terms. More than a biotic inventory of microorganisms is needed to develop a complete understanding of the organization and interactions within the community. Individual microorganisms display such strong interactions that new capabilities are needed to understand recalcitrance in comprehensive and integrated way. Therefore, a “systems biology” approach is needed to envision how biomass decomposition functions as a complete set of intersecting processes. This requires that we also understand the biochemical reactions and catalysts involved in the deconstruction of biomass. Clearly, no technique alone can provide the broad range of information needed to understand community structure and system function.

Plant biomass is a chemically diverse substrate varying in composition, but predominated by complex and interactive polymers of cellulose, hemicellulose, and lignin. Evolution has developed plants that are naturally recalcitrant to degradation by microbial communities. Major contributing biochemical features to the recalcitrance of terrestrial plants not shared by bryophytes and earlier plants are the lipid materials of the cuticle and its wax, and the diversity of phenolics in lignin and flavinoid compounds (4). The arrangement and density of the vascular bundles, the relative amount of sclerenchymatous (thick wall cell) tissue, the structural variation, and complexity of cell wall constituents also contribute to recalcitrance. The result is that the natural plant biomass decay involves diverse groups of heterotrophic bacterial and fungal communities that degrade and metabolize plant material. In short, Nature’s answer to the innate chemical and structural complexity found in plants is greater diversity and synergistic interactions. The full extent of this diversity is still up to debate.

Carbon turnover in the terrestrial biosphere occurs chiefly in the soil from complex and yet unclassified microbial communities. Turnover rates vary dramatically depending on environmental conditions such as temperature, water availability, inorganic nutrients, pH, organic carbon input such as plant exudates, biomass composition, and the presence of microorganisms producing hydrolytic enzymes such as cellulases and hemicellulases. The exact mechanisms of degradation carried out by complex interactions of members of biomass-degrading community remain elusive. The arrival of new biotechnology tools such as ribosomal rRNA gene sequencing, comparative metagenomics, transcriptome, and secretome analysis are allowing for habitat-specific fingerprinting of microbial communities. Genomic-based studies are providing vast amounts of sequence data from various environmental samples and have led to new insights into microbial populations. The improvement

of sequencing technologies has made metagenome shotgun sequencing of an environmental sample feasible; however, most environmental communities are far too complex to be fully sequenced in this manner.

Metagenomic analysis of representative bacterial assemblages is beginning to provide a knowledge base and a source of genetic material for further studies on the ecology of lignocellulose degradation and biotechnology applications. One obvious limit of environmental metagenomic sequencing is the sheer diversity of the microbial communities that populate rich environments, which calls for large sequencing efforts, to assemble long DNA sequence data. This complexity can lead to bias through overrepresentation of only a few common genomes, and because of the difficulty of assembling long sequences for analysis. Also of significance is the increasing number of published genomes of biomass-degrading organisms. Genomics research is not only providing phylogenetic information about microbial populations but in conjunction with new biochemical tools is generating a broader understanding of the complete biophysical activity of the community. Together these new tools are rapidly advancing our understanding of the principles that underlie microbial biomass degradation and carbon cycling in nature. The knowledge gained from these studies is a stepping-stone to the development of optimized enzymes and microorganisms for the production of commodities such as ethanol and hydrogen from biomass.

15.2 Defining biomass decay communities

Microorganisms involved in the decay of plant biomass use an array of biochemical processes to mediate conversion. They can be categorized either genetically using molecular biology tools and/or biochemically, with emphasis on the types of hydrolytic enzymes they produce. The biodegradation of complex polysaccharides is a multistep, hierarchical process. It is hierarchical in two senses. In general, removal of sugar modifications makes subsequent attack of main chain-degrading enzymes much more effective. Secondly, in nature there is a hierarchy of polymer degradation, from the least to the most recalcitrant compounds. During the early stages, the rate of decay is determined by the accessibility of the more easily decomposed carbohydrates. In general, pectins are degraded first, followed by hemicellulose, and then finally cellulose and lignin. This results in a succession of microorganisms that are able to assimilate specific substrates.

Hydrolytic enzymes capable of breaking down cellulose and hemicellulose (complex polymers of arabinose, mannose, glucose, and xylose) are widespread and are produced by a wide range of microorganism that include fungi, actinobacteria, Clostridiaceae, and members of the α -, β -, and γ -proteobacteria. The efficient conversion of plant polysaccharides is achieved by an array of enzymes with varying but synergistic specificities and activities. The diversity of these enzymes in nature is illustrated by the growing number of entries of hydrolytic enzymes in glycosyl hydrolase-specific databases such as the Carbohydrate-Active enzymes (CAZy) database (5). With the increasing number of sequenced genomes available it was recently estimated that over 12 000 glycosyltransferase and glycoside hydrolase open reading frames will have been added to the database during 2006 (6).

A second and more narrowly defined functional group includes the white rot Basidiomycota and xylariaceous Ascomycota, which are the primary agents of lignin degradation by secretion of nonspecific, extracellular enzymes that modify the lignin macromolecule. Lignin

is a polyphenolic heteropolymer that is interconnected by stable ether and carbon–carbon bonds making it extremely resistant to microbial attack. Lignin degradation is carried out by the action of lignin modifying enzymes that are produced primarily by the white rot fungi. Lignin breakdown is an obligate aerobic reaction carried out by oxidative ligninolytic enzymes that remove various functional groups, side chains, and aromatic rings randomly from the lignin macromolecule. The initial steps often involve O_2 - or H_2O_2 -dependent, lignin peroxidase, Mn peroxidase, or laccase. Frequently, flavin-dependent enzymes supply H_2O_2 to the lignin peroxidase or Mn peroxidase (7, 8). The lignin polymer is reduced to smaller, low molecular weight fragments that are then available for further decomposition by either fungi or bacteria.

Historically, microbiologists have defined decay communities based on individual isolated bacteria and on the basis of their physiologic and nutrition requirements. In general, classical isolation techniques require substantial knowledge of the organism's habitat so that it can be reproduced and one population can be enriched over others. Within a small particle of biomass several different microenvironments can exist differing both chemically and physically. Physiochemical conditions, such as oxygen concentration change both temporally and spatially during decay. In many cases, individual bacteria cannot be isolated due to obligate interdependence upon other bacteria for growth that must be supplied using co-cultures of helper bacteria. Microorganisms may also be slow-growing, or obligate anaerobes requiring patience and tenacity to isolate and characterize. With the advent of new molecular tools for describing microbial communities, it has become clear that bacteria grown isolated in "pure culture" underrepresent the microbial diversity in most environments and are generally only a minor component.

15.2.1 Fungi identified with plant biomass

Fungal species play a critical role in plant biomass decay and can be divided into the following three categories: (a) saprophytic fungi, which prefer dead and decaying material; (b) parasitic fungi, which prefer colonizing the biomass of the living host; and (c) mycorrhizal fungi, which form partnerships with specific plant species, mostly with living trees. Most fungi are saprophytic and are found in nature growing on dead organic matter. They are effective at secreting enzymes that degrade large polymers such as cellulose, hemicellulose, lignin, pectin, starch, and protein found in the organic matter to release nutrients that can be taken up and used by the fungi. The filamentous fungi are widespread in nature and for most part are obligate aerobes. Typically, fungi colonize substrates with tubular, branching hyphae that collectively form the mycelium. The hyphae cell walls composed of β -glucans and chitin embedded in a matrix that includes β -glucans and glycoproteins. As shown in Figure 15.1, growth is by extension of the hyphal tips that enables the mycelium to spread and penetrate throughout the substrate allowing new areas to be colonized as the nutrients are used from the earlier colonized substrate. Extensive communication and nutrient transfer is possible through the hyphae and between different regions of the mycelium. Older parts of the mycelium may die after nutrients in the substrate are exhausted and material from these areas can be transported to younger regions of the mycelium. Many ascomycetes and basidiomycetes produce a group of small (70–120 amino acids) amphiphilic proteins that promote attachment of hyphae to hydrophobic surfaces (9, 10). Hydrophobins are

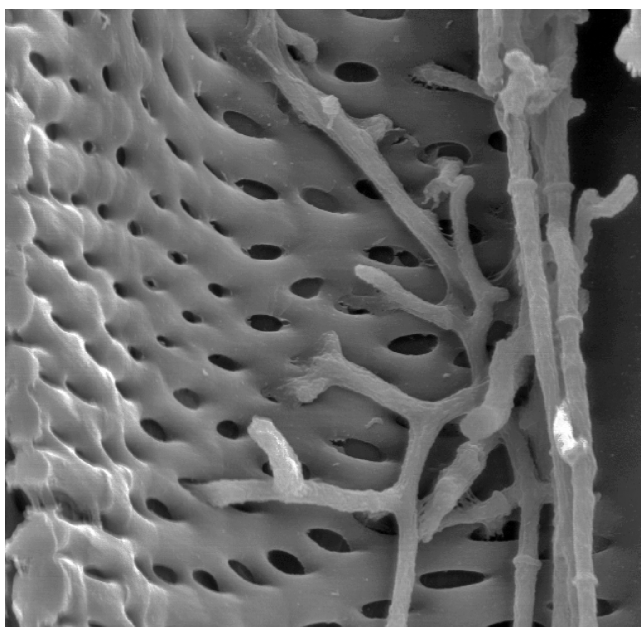


Figure 15.1 Scanning electron micrograph of corn stem inoculated and colonized by the cellulase producing fungus *Trichoderma reesei*. Fungal mycelia are attached to and can be seen penetrating anatomical structures like the apoplast and pores/pits. Image was generated in the NREL Biomass Surface Characterization Laboratory and provided by Todd Vinzant.

self-assembling proteins that have been demonstrated to function at hydrophilic–hydrophobic interfaces to form films and to function as surfactants (10). Hydrophobins have been proven to lower the surface energy of water allowing the fungal hyphae to penetrate the air–water interface and grow up into the air (11, 12). Because of their adhesive and surface activity, hydrophobins are an interesting feature of biomass colonization.

Producing plant cell wall-degrading enzymes is widespread in fungi and has been described in the anaerobic fungi found in ruminants (13) and all other subdivisions of aerobic fungi including members of the Zygomycetes, Ascomycetes, Basidiomycetes, and Deuteromycetes. A major ecological activity of many fungi, especially members of the Basidiomycetes, is the decomposition of lignocellulosic materials such as wood and other plant material. Fungal breakdown of biomass occurs by succession where the species of one fungal community alters the substrate enough to allow other species to become established and colonize (14). Two types of wood rot are known: brown rot, in which the cellulose is preferentially used and the lignin is not metabolized and white rot where both cellulose and lignin is used. White rot, brown rot, and soft rot fungi are recognized among those that colonize dead wood. Soft rot fungi degrade cellulose and hemicellulose in the wood under conditions of high moisture content leaving the wood soft or spongy without being fully degraded. Soft rot has also been reported to occur in dry environments and under other extreme conditions (15). Soft rot decay of wood reportedly occurs at historic sites in Antarctica where extreme environmental conditions inhibit wood decay from other fungi

(15). Soft rot is caused primarily by fungi classified in the phylum Ascomycota with the decay characterized by cavities that form within the biomass structure. Soft rot fungi can also cause a progressive degradation of secondary cell walls that can be completely degraded except for the middle lamella between the cells (15). The white rot fungi exhibit a large amount of diversity, but are generally members of the Basidiomycetes or other higher fungi, and produce enzymes that are degraded lignin in addition to cellulases and hemicellulases (16). The hyphae of white rot fungi rapidly colonize wood by growing within the lumen of the cells and degrading the cell walls. Although they can degrade the substrate, white areas of the wood remain where the lignin and hemicellulose have been removed ahead of the cellulose. Under conditions where the biomass is moist members of the Basidiomycota will rapidly degrade cellulose and lignin by generating oxidants such as hydroxyl radicals.

Parasitic fungi are the second largest group, of whose members do considerable damage to growing plants. Indeed, root diseases and mycorrhizal systems have a similarity regarding parasitism. Root pathogens, such as *Rhizoctonia*, *Fusarium*, *Verticillium*, *Sclerotinia*, and *Pythium* are stimulated by roots. These fungi progressively invade and colonize the meristematic root tissues and ultimately cause necrosis.

The rhizosphere is the region immediately outside the root that generally has more microbial activity than the surrounding soil. Mycorrhiza refers to the symbiotic association between plant roots and fungi classified as ectomycorrhizae where the fungi form an extensive sheath around the root, and arbuscular mycorrhizae where the fungal mycelium becomes embedded with root tissue. Mycorrhizal associations are widespread and are found on most plants in diverse environments. Although the production of hydrolytic enzymes has been described in many mycorrhizal fungi, most do not use cellulose for metabolism but obtain carbon from root secretions (17). Mycorrhizal fungi have been demonstrated to have weak cellulase and endopolygalacturonase activities (17). Hohnjec *et al.* described the presence of three different endoglucanases with different preferences for sugar bonds and four different pectinolytic or polygalacturonate-degrading enzymes in *G. mosseae* and *G. intraradices*-colonized Medicago roots (18). Cellulase, pectinase, and xyloglucanase activities have been found in the external mycelium of arbuscular mycorrhizas (19). The production of these hydrolytic enzymes is thought to allow one for the modification of the extracellular matrix of the root and allow fungal colonization (18).

15.2.2 *Bacteria identified with plant biomass*

Bacteria are by far the most numerically abundant and taxonomically diverse microorganisms in bulk soil and are found both free-living and attached to the surface of soil particles. Many soil bacteria also interact with the roots of plants within the rhizosphere. Many of these bacteria have also been shown to possess hydrolytic enzymes that enable them to colonize and degrade plant biomass. They are widely distributed across phylogenetic groups that include diverse functional groupings. Many are native to soil, but have also been isolated from such diverse environments as hot springs, rumen contents, compost piles, termite guts as well as various other sites. The biochemical characteristics and physiology of biomass-degrading prokaryotes vary according to the environmental niche where they are found. Due, in part, to the diversity of prokaryotic microorganisms that produce cellulolytic enzymes and the promise of newer and better high-specific activity enzymes for industry, (Diversa

is now called Verenum Corporation) these systems remain the focus of considerable study. Companies like Diversa Corporation in San Diego, California, collect DNA from hydrothermal and other habitats worldwide, and then screen-extracted genomic DNA for the ability to produce useful enzymes for biomass conversion applications. Diversa's business platform since 1995 has been to discover novel enzyme by creating gene libraries from bioprospecting globally diverse environments. Their ultimate goal is to develop enzymes capable of surviving at extreme process conditions for industry. Another company, Dyadic International, Inc, uses an integrated high-throughput technology platform targeted specifically at the discovery, expression, and modification of both prokaryotic and eukaryotic genes.

For microorganisms to degrade and metabolize the insoluble polysaccharides found in plant cell walls they must produce several extracellular hydrolytic enzymes. These enzymes can be secreted and act free in solution, or are cell associated. Adherence or colonization to insoluble substrates by microorganisms is common. Attachment can also be a factor in the control of enzyme expression and can be a precondition to the production of hydrolytic enzymes. Commonly, microorganisms act synergistically to enable efficient conversion of the substrate by the concerted action of several species. Synergism may involve the production of bacterial communication signals (quorum sensing) such as acylated homoserine lactones (AHLs) to establish these communities (20, 21). Quorum sensing has also been reported as a regulatory system for the control of extracellular enzyme synthesis in phytopathogenic bacteria, and in nitrogen-fixing rhizobia (22, 23). Bacterial communication signals may also allow the cells to regulate secretion of hydrolytic enzymes to reduce losses because of diffusion.

Because cellulose is a large insoluble polymer its use first requires binding of the enzymes either as a binary enzyme-substrate complex or as a ternary enzyme-substrate-microbe complex. Adhesion is most pronounced in the Gram-positive, thermophilic anaerobes such as *Clostridium thermocellum* or *Clostridium cellulolyticum*, which secretes an active and thermostable high molecular weight cellulase complexes (cellulosome) responsible for degrading crystalline cellulose (24, 25). Cellulosomes contain at least 30 polypeptides, most of the enzymes are endoglucanases (EC:3.2.1.4), but there are also some xylanases (EC:3.2.1.8), β -glucosidases (EC:3.2.1.21), and endo- β -1,3-1,4-glucanases (EC:3.2.1.73). Hydrolysis of cellulose by cellulosome producing organisms is dependent on adherence of the organism to the substrate through specific cellulose-binding proteins.

At least 46 unique bacterial producers of cellulases have been identified from many aerobic bacterial systems, including species within the genera *Acidothermus*, *Bacillus*, *Cellulomonas*, *Cellvibrio*, *Cytophaga*, *Microbispora*, *Pseudomonas*, and *Thermobifida* (26). Anaerobic bacteria identified as biomass degraders include members of the genera *Acetivibrio*, *Bacteroides*, *Clostridium*, *Micromonospora*, and *Ruminococcus*. Bacteria that decay biomass have been isolated from temperature extremes that include psychrophilic, mesophilic, thermophilic, and hyperthermophilic conditions. The actinobacteria, for example, are widespread microbial components of terrestrial and aquatic communities that have been demonstrated to play key roles in biomass turnover and nitrogen dynamics. Like fungi the actinobacteria have the ability to penetrate lignocellulosic biomass which gives them the ability of secreting hydrolytic enzymes in confined cavities allowing for higher concentrations of free enzymes.

Recently sequenced genomes representative of the actinobacteria include organisms that are widespread in plant rhizospheres, plant tissues, and compost, including *Streptomyces*

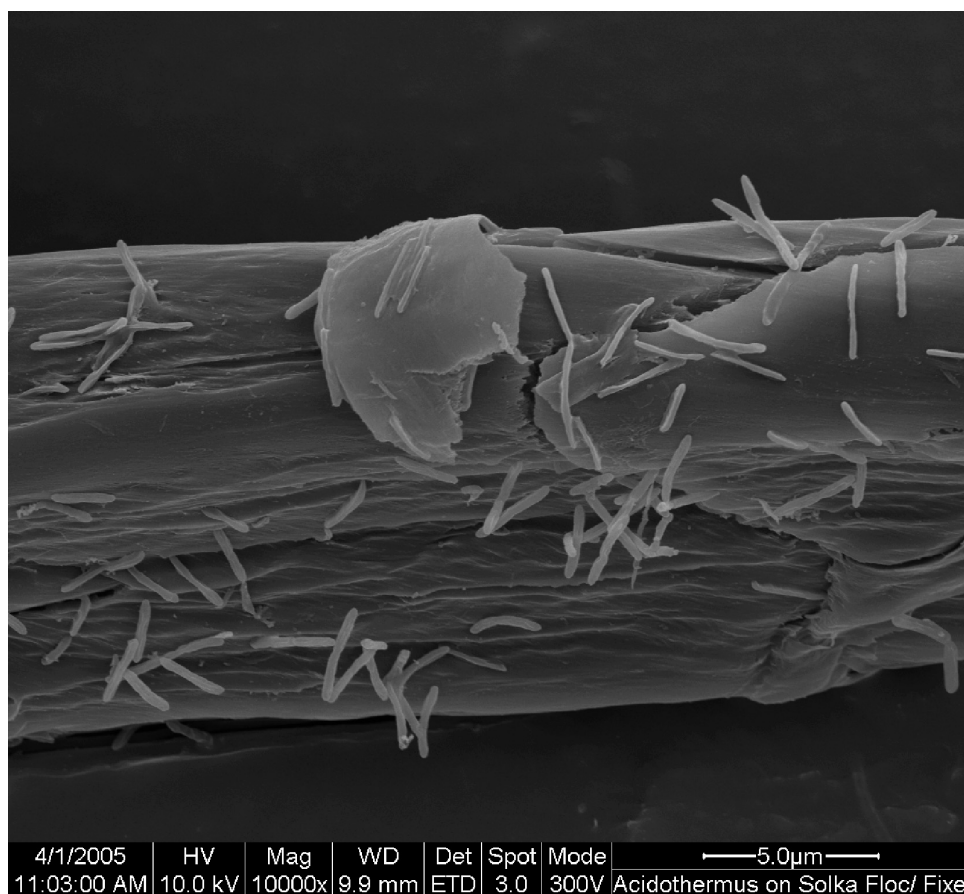


Figure 15.2 Scanning electron micrograph of a cellulose microfibril that is colonized by the cellulolytic actinobacterium *Acidothermus cellulolyticus*. Image was generated in the NREL Biomass Surface Characterization Laboratory and provided by Todd Vinzant.

spp., *Thermobifida fusca*, and *Leifsonia xyli*. A close relative to *Acidothermus cellulolyticus* and *Frankia*, *Kineococcus radiotolerans* is a drought-tolerant soil microbe that is also radiation tolerant, while *Rubrobacter* is a thermophilic genus containing radiotolerant species. The genomes of all these taxa contain cellulose- and other plant biomass-degrading metabolic pathways. The high-GC actinobacteria represent a large phylogenetic group and serve as a reservoir of organic carbon and nitrogen cycling capabilities. Many species of actinobacteria have adapted to diverse and often extreme environmental conditions. For example, Actinobacteria have been identified as the major group of cellulolytic organisms in ecosystems such as acid *Sphagnum* peat bogs (27).

The genomic sequence of *A. cellulolyticus*, a major cellulose degrader (see also Figure 15.2) originally isolated from the hot springs samples has recently been compiled and published by the Department of Energy Joint Genome Institute, accession NC_008578. *A. cellulolyticus* was originally isolated from submerged woody debris in a Yellowstone National Park hot

spring, because of its efficient cellulose degradation (28). The organism was isolated using selective culturing methods of mud and decaying wood samples from thermal features in and around the Norris Geyser Basin region of the park. The microorganism was found to grow at 37–65°C with an optimum of 55°C. A type strain was selected and deposited in the American Type Culture Collection and is part of their National Park Service special collection (<http://www.atcc.org/SpecialCollection/NPS.cfm>). The genome contains a suite of glycosyl hydrolases responsible for cellulose degradation that have already been characterized as thermostable (29) as well as a xylanase and additional multifunctional hydrolytic activities.

The genome of the closely related thermophile, *Thermobifida fusca*, has also been completed. The *T. fusca* genome shares a high number of gene homologs with *A. cellulolyticus*. Also of interest are members of the genus *Frankia*, which are filamentous actinobacteria that form nitrogen-fixing root nodules on woody trees and shrubs in a symbiosis termed “actinorhizal.” Genome sequence data are now available for three strains of *Frankia*, an ecologically important nitrogen-fixing root nodule symbiont which is the closest phylogenetic relative to *A. cellulolyticus*. The sequenced *Frankia* genomes contain homologs for cellulases, xylanases, peroxidases, and aromatic degradative enzymes, as well as genes for secondary metabolite biosynthesis. Similar metabolic capabilities have been identified in the genomes of *Streptomyces* spp., suggesting that actinobacteria may be a rich source of enzymes for bioconversion. Members of the *Frankia* have been found to inhabit root nodules, rhizosphere (termed actinorhizal), and the soil as a saprophyte. *Frankia* – actinorhizal plant symbiosis has been reported within a range of actinorhizal plants scattered among eight families consisting of over 200 species of angiosperms (30, 31).

Thermobifida fusca is a moderate thermophilic soil actinomycete (growth temperature ranging from minimal 30°C to maximal 55°C) that is a major degrader of plant cell walls in heated organic materials such as compost heaps, rotting hay, manure piles, or mushroom growth medium. It produces spores that can be allergenic and causes a condition called farmer’s lung. Its extracellular enzymes, including cellulases, have been studied extensively because of their thermostability, broad pH range (32–37), and high-specific activity. It degrades all major plant cell wall polymers except lignin and pectin and can grow on most simple sugars and carboxylic acids. Its genome has been sequenced and closed by the JGI; it has a size of 3.4 Mb and encodes seven known cellulase genes, all which have been expressed and characterized. Four of the cellulases are endocellulases (Cel5A, Cel5B, Cel6A, Cel9B), one is an exocellulase attacking the nonreducing ends of cellulose molecules (Cel6B), one is an exocellulase attacking the reducing ends of cellulose molecules (Cel48A), and one is a new class of cellulase, a processive endocellulase, which have now been found in several cellulolytic bacteria. All the cellulases contain a family 2 cellulose-binding module (CBM) attached by a linker peptide. There are many hemicellulose genes in *T. fusca* and five of them have been expressed and characterized: Xyl11A, Xyl10A, Xyl10B, XG74A, and Gn81A.

There have also been extensive studies of the regulation of *T. fusca* cellulases that have shown that their synthesis is induced by cellobiose and repressed by any good carbon source. All the six-cellulase genes encoding cellulases that are regulated by cellobiose contain at least one copy of a 14-base inverted repeat sequence upstream of their start codon. This sequence is the binding site for a lacI family regulatory protein, CelR. CelR has been expressed and characterized in *E. coli*. Its binding to DNA containing the 14 base sequence

is inhibited by cellobiose. There are a number of copies of this 14 base sequence in *T. fusca* DNA, including one copy upstream of an operon that contains a β -glucosidase gene and genes for a binding protein transport system. The two exocellulases make up about 75% of the total cellulase produced by induced *T. fusca* and both exocellulase genes contain a second copy of the CelR-binding site upstream of their start codon. It is not clear if this regulatory site is responsible for their higher transcription rate. It is also not known if CelR plays a role in carbon source repression.

The genome of *Cytophaga hutchinsonii*, another eubacterial cellulose degrader, has also recently been sequenced. *C. hutchinsonii*, an aerobic Gram-negative bacterium commonly found in soil that has been shown to rapidly digest crystalline cellulose (38). Molecular analysis of cellulose degradation by *C. hutchinsonii* is now feasible, since techniques for the genetic manipulation of this organism have recently been developed (39). *C. hutchinsonii* belongs to the Cytophaga–Flavobacterium branch of the eubacterial phylogenetic tree. Members of this group are widely distributed in many environments and have the ability to move rapidly over surfaces by gliding motility (40, 41). Gliding motility is thought to be important in allowing *C. hutchinsonii* to colonize its insoluble growth substrate. The genome of *C. hutchinsonii* was sequenced by the JGI and has been shown to differ from most known cellulolytic microorganisms in that none of its cellulase genes encode processive cellulases and few of them encode a CBM. These results provide strong evidence that *C. hutchinsonii* does not use either of the two known mechanisms for cellulose degradation: secretion of a set of individual synergistic cellulases containing CBMs or production of cellulosomes, since processive cellulases and CBMs are key for both mechanisms. Determining the detailed mechanism of cellulose degradation by *C. hutchinsonii* is a major unsolved problem in plant cell wall degradation and this pathway might provide new proteins that improve the rate of cellulose degradation.

Recently, an interesting group of marine microorganisms, the marine complex polysaccharide (CP)-degrading bacteria from the *Microbulbifer*, *Teredinibacter*, and *Saccharophagus* group have been described. This group of organisms produces an array of enzymes to degrade complex polysaccharides including cellulose (42). The genome of the marine bacterium, *Saccharophagus degradans*, has been completed and reportedly contains more than 180 open reading frames that encode carbohydrate-degrading proteins (43). *S. degradans* is a pleomorphic, Gram-negative, aerobic member of the γ -proteobacterium isolated from decaying salt marsh cord grass.

C. phytofermentans is a recently discovered member of the order *Clostridiales*. Its genome is presently being sequenced by the JGI. It was found in forest soil and has a broad growth substrate range and is able to rapidly degrade and ferment several plant polymers, including cellulose, pectin, starch, and xylan (44). A remarkable property of cellulose-fermenting *C. phytofermentans* cultures is the production of high concentrations of ethanol, typically more than twice the concentration produced by other cellulolytic clostridia, and hydrogen.

15.3 Interactions between saprophytic fungi and bacteria

Microorganisms associated with the rhizosphere have been demonstrated to increase root exudation through production of plant hormones and by physically damaging the roots (45). These actions create a nutrient-rich rhizosphere zone that is naturally colonized

by many beneficial or sometimes pathogenic microorganisms. Bacteria and fungi have a considerable impact on plant growth, development, and productivity. The numerous interactions between bacteria, fungi, and roots may have beneficial, harmful, or neutral effects on the plant, the outcome being dependent on the type of symbiont interaction and the soil conditions (46). "Helper bacteria" may promote the formation of fungal communities in decaying biomass litter (47). Little is known about the ecology of these helper bacteria; however, an analogy to the role played by bacteria present in mycorrhizal fungi can be made (47). In the rhizosphere of plants, and perhaps in the decaying biomass pile, fungi are always accompanied by another important group of microorganisms, the bacteria which also prosper in the organic-rich environment (mostly sugars, amino acids, and organic acids) released from the roots and mycorrhizal fungi. Among them, there is a large category of growth-promoting rhizobacteria that influence plant growth directly or indirectly by releasing a variety of compounds, from mineral nutrients, to phytohormones, and antimicrobial compounds. Rhizobacteria have been demonstrated to promote plant growth directly through production of plant hormones such as auxins (48), gibberellins (49), and ethylene (46). Production of indole-3-ethanol or indole-3-acetic acid (IAA), compounds belonging to the auxins, have been reported for several bacterial genera, such as *Frankia* (50–52), *Klebsiella* and *Enterobacter* (53), and *Bacillus* (49). Most importantly, for rhizobacteria to act beneficially, they must be able to efficiently colonize and multiply in the plant rhizosphere.

Electron microscopic studies and molecular methods revealed large bacterial populations associated with mycorrhizal roots and extraradical hyphae (54–56). In some cases, bacterial endosymbionts were discovered in fungal hyphae (55, 57). The establishment of mycorrhizas on roots is affected by the microbial populations of the rhizosphere, and especially by some bacteria, which can have either a positive or a negative effect on mycorrhiza formation. Garbaye (47, 58) defined a new bacterial category, the mycorrhization helper bacteria that strongly promoted ectomycorrhiza formation. Frey-Klett and coworkers (59) suggested that these helper bacteria stimulate the growth of fungal mycelia, thus increasing the probability of a root-mycelium encounter. Some mycorrhiza helper bacteria were noncultivable, and were identified only by molecular ecological methods.

15.4 Characterization of microbial communities that degrade biomass

Microorganisms associated with upper soil and decaying biomass are comprised of a relatively small group of microorganisms that can be grown *in vitro* and a larger majority that cannot be grown. The most extensive zone of microbial growth occurs on surfaces of biomass, usually within the rhizosphere region of soil. The most studied biomass degrading microbial communities are those involved composting, wet wood, estuaries, and forest floor (60–63). Historically, the ecology of these sites has been investigated by examining those organisms that could be isolated and grown in the laboratory or by measuring the enzymes they produce. Most reports describing the population dynamics have been done by traditional culture and phenotyping methods (64), through the use of systems such as the BIOLOG method (65, 66), by the measurement of phospholipid fatty acid patterns in soil or litter samples (67, 68), and by extracting and monitoring enzyme activities (69). It

has also been reported that augmentation of a biomass with specific enzymes, primarily cellulases, stimulates decomposition (70), although the mechanisms at the community or molecular level are not known.

15.4.1 Biochemical approaches to define biomass degrading communities

Extracellular enzymes catalyze the initial rate-limiting step of decomposition and are the primary means by which microbes degrade complex biomass into smaller molecules that can be assimilated. Therefore, it is reasonable to assume that determining the amount and type of hydrolytic activities associated with biomass can be an indicator of the hydrolytic potential of a given microbial community. Defining biomass degrading communities through the application of traditional biochemistry tools combined with new proteomic technologies are advancing our understanding of how microorganisms attack biomass as well as our understanding of natural enzyme diversity. Proteomic analysis can be used to identify and validate hydrolytic enzyme targets and profile protein expression patterns in complex communities (71). By using two-dimensional PAGE it is possible to fingerprint the secreted proteins of cellulose-degrading microorganisms and obtain sufficient sequence information for cloning. Surveying the proteome of natural microbial communities can lead to the discovery and analysis of more diverse hydrolytic enzymes that can break down cellulose, hemicellulose, and lignin. New classes of ligninases and hemicellulases will likely be identified, their mechanisms of action understood, and their performance refined to allow introduction of enzymatic pretreatment that will free cellulose microfibrils for enzymatic saccharification (breakdown to sugars).

The traditional biochemical approach to understanding microbial biomass utilization is through kinetic assays designed to quantitatively determine the presence or absence of hydrolytic enzymes. Hydrolytic enzymes can be detected using either natural substrates such as cellulose or xylan by measuring sugar release or by using synthetic substrates that contain easily detected chromophores such as *p*-nitrophenol or 4-methyl umbelliferone (69). The detection of 4-methyl umbelliferone can be used as a sensitive, quantitative assay for endoglucanase or other enzymes that cleave substrates linked to 4-methyl umbelliferone (72). However, how tightly the enzymes are bound to the substrate or are associated with the microorganism is a limitation of direct enzymatic assays. A range of soil enzyme assays was developed by Lynch and coworkers as alternatives to population measurements (46). These included assays for determining chitinase, *N*-acetyl glucosaminidase, β -glucosidase, β -galactosidase, acid phosphatase, alkaline phosphatase, phosphodiesterase, aryl sulfatase, and urease activities from small soil samples. Soil enzyme activities, therefore, can index changes in the microbial functioning in soil, and there is ample evidence in the literature of the importance of glycosyl hydrolases, and proteases to the soil's performance (69, 73–77).

Monitoring many proteins simultaneously in a complex system can be best accomplished for many hundreds of protein species across a large number of samples using modern technology for two-dimensional differential gel electrophoresis in combination with sophisticated statistical algorithms for data analysis. Two-dimensional gel electrophoresis is capable of resolving several hundreds to several thousands of proteins on a single gel (78). This method utilizes independent properties of proteins (i.e., isoelectric point and molecular

mass) to resolve proteins present in a biological extract in two dimensions. Methods for two-dimensional differential gel electrophoresis have been greatly improved over the last several years to enable quantitative analysis of relative protein abundance among a set of samples. The first major improvement involves derivatization of proteins in samples with spectrally distinct, covalently coupled, charge- and mass-balanced fluorescent dyes (79). These highly fluorescent tags allow for extremely sensitive detection limits [i.e., 1 ng; (80)] and a broad linear response range [i.e., 3–4 orders of magnitude; (81)]. Derivatization of two different protein extracts with spectrally distinct fluorescent tags enables multiplexing of two unknown samples on gels, which eliminates any question about which protein in one sample is co-migrating with what spot on another gel. A second major improvement of the experimental design for two-dimensional differential gel electrophoresis is the inclusion of an internal standard labeled with a third, spectroscopically resolvable, fluorescent dye (82). This enables normalization of the data from a series of gels, which permits statistically valid comparison of protein amounts across a series of gels. Because of this improved experimental design, it is now possible to detect changes in absolute protein abundance on the order of 10% with 95% confidence.

15.4.2 Molecular approaches for defining biomass-degrading communities

The challenge of studying the ecology of biomass-degrading communities has always been the ability to accurately determine a representative picture of the true diversity of the microbial consortia, and the biochemical processes. Advances in molecular biology have led to development of culture-independent methods for describing microbial communities based on analysis of DNA extracted directly from natural populations thereby circumventing the need to isolate and culture bacteria for phylogenetic analysis. The recent surge of research in molecular microbial ecology has provided evidence for the existence of many novel types of microorganisms in the environment, in numbers, and varieties far greater than cultivated in the laboratory, which probably comprise less than 1% of all microorganisms (33). Additional corroboration comes from estimates of DNA complexity and the discovery of many unique bacterial 16S rRNA gene sequences from numerous environmental sources (2, 35, 83). One approach to classifying complex communities is to use “marker” genes as phylogenetic anchors for identification of source microorganisms. Ribosomal RNAs are highly conserved and are commonly used to determine phylogenetic relationships between organisms. Other marker genes can be used including *recA* and *rpoB* genes, however, the largest available database is for 16S rRNA. Another approach is to use gene-centric analysis of large datasets that focus on the identification and taxonomic characterization of genes important to the overall community function.

DNA-based fingerprinting techniques developed to characterize and compare whole genomes of organisms include amplified fragment length polymorphism (84), terminal restriction fragment length polymorphism (68), denaturing gradient gel electrophoresis (70), amplified rRNA gene restriction analysis (85), restriction landmark genome scanning (86), and automated ribosomal intergenic spacer analysis (87). Dominate members of biomass-degrading communities can be determined using methods such as terminal restriction fragment length polymorphism (T-RFLP) and denaturing gradient gel electrophoresis (DGGE).

15.4.2.1 Comparative sequencing and quantification of rRNA

Measuring microbial diversity typically involves sequencing individual 16S rDNA gene sequences, to obtain species-level resolution. The power of the approach comes from the ability to amplify, clone, and analyze homologous regions of 16S rDNA from small amounts of sample DNA. Some studies focus on only a small portion of the 16S rDNA, while others survey the entire gene sequence. Sequencing the entire 16S rDNA enhances the accuracy of estimating the species variability of microbial communities, and is preferred when broad comparative determinations are preformed.

The question of whether two populations of microorganisms have different numbers of species and the level of genetic diversity can be answered by comparing the relationships and degrees of divergence among sequences. Depending on amplification and cloning conditions the technique can overrepresented some species while underrepresenting others. There is also a relatively small potential for error from sequence variation due to PCR replication errors. Variations of the 16S approach include denaturing gradient gel electrophoresis and in situ hybridization.

Bacterial populations, including soil communities, have been characterized using rRNA intergenic spacer analysis (RISA) which determines the variability in the length of the intergenic spacer between the small (16S) and large (23S) ribosomal subunits. The method has been automated (ARISA), and the sensitivity increased by the use of fluorescence-tagged oligonucleotide primers for PCR amplification and for subsequent electrophoresis in an automated capillary electrophoresis system. Fungal ARISA makes use of the length polymorphism of the nuclear ribosomal DNA (rDNA) region that contains the two internal transcribed spacers (ITS) and the 5.8S rRNA gene. Comparative sequencing and quantification of rRNA genes using universal and phylogenetically specific primers has become an established method for detecting and characterizing subgroups of prokaryotic and eukaryotic microorganisms. For instance, constructed 16S and 18S rRNA gene libraries have been constructed to examine the effects of elevated CO₂ levels on the composition of microbial communities associated with the rhizosphere of trembling aspen (88). However, a thorough analysis of complex microbial communities using rRNA gene-based libraries requires a huge sequencing effort. To overcome this sequencing limitation, several rapid high-throughput DNA-based molecular methods for rRNA gene-based analysis of microbial communities have been developed. These methods, including terminal restriction fragment length polymorphism (T-RFLP) and denaturing gradient gel electrophoresis (DGGE), provide a DNA fingerprint of the microbial community present in the sample (89–92). While T-RFLP and DDGE are sensitive methods for differentiating between microbial communities, these methodologies do not provide actual DNA sequence information unless the resolved bands are recovered and sequenced (93).

15.4.2.2 Metagenome analysis

Collectively, the genomes of the total microbiota found in nature, referred to as the metagenome (37), contains vastly more genetic information than is contained in the cultivable subset. However, the genetic complexity of a microbial community at a specific site is influenced by many environmental factors. Re-association of total community DNA extracted from different environmental distinct sites has revealed that the community genome size can equal that of 6000–10 000 *Escherichia coli* genomes in unperturbed organic

soil, but only 350–1500 genomes in arable or heavy metal-polluted soils (83, 94). These estimates are conservative, since genomes representing rare and unrecovered microorganisms were probably not included in the analysis. As expected, Torsvik and Ovreas (2) could recover less than 40 genomes by culturing methods which emphasizes the need for development of novel methods and approaches to provide new insight into the relationship between phylogenetic and functional diversity of these communities as ecosystems.

DNA sequencing continues to be one of the most important platforms for the study of biological systems. With the development of improved sequencing technologies that enhance the speed, sensitivity and throughput, it has become feasible to sequence the entire metagenome of an environmental sample (95). Culture-independent genomic analysis of microbial communities using metagenomics is revealing that soil and ocean environments are more genetically and potentially more biochemically diverse than previously thought (96). This involves the cloning and analysis of large genomic DNA fragments isolated from a mixed community. The metagenomic library can then be screened for functional or taxonomic genes of interest or sequenced by shotgun sequencing. Most environments contain communities far too complex for it to be possible to sequence a complete metagenome, and even the simple communities contain micro-heterogeneity that makes most genome reconstructions simplified versions of reality. Reconstruction of community metagenomes was initially pursued for viral communities in the ocean and human feces (97–99) and has since been attempted in an acid mine drainage (AMD) biofilm (100) and the Sargasso Sea (101). The AMD biofilm community was ideal for complete metagenome sequencing because 16S rRNA gene sequencing indicated that there were three bacterial and three archaeal species in the biofilm. Marine communities contain far greater species richness, on the order of 100–200 species per milliliter of water (102, 103), making the sequencing and assembly effort considerably more difficult. Further out on the continuum of biological complexity is soil, with an estimated species richness on the order of 4000 species per gram of soil (35, 102, 103). Sequencing the soil metagenome requires faster and less expensive sequencing technology than currently available.

Recently, we initiated in collaboration with the JGI the sequencing of the metagenome of a microbial community actively decaying poplar biomass under anaerobic conditions. The predominance of microbial enumeration in the biomass pile is represented by this large anaerobic core zone. In addition to some cellulolytic fungi, bacteria of the order *Clostridiales*, many of which have strong cellulolytic activities, were found to dominate this specific microbial community. The estimated composition and the distribution of bacterial members of this community were determined based on 16S rRNA gene sequencing (S. Taghavi and D. van der Lelie, unpublished). It should be noted that we were able to cultivate several members of this community and characterize their cellulolytic activities. Interestingly, none of these cultivable species represented the dominant community members, stressing the importance to use a cultivation-independent approach to characterize the composition and metabolic potential of this complex microbial community.

15.4.2.3 Expression profiling

An emerging alternative approach is to use oligonucleotide or cDNA microarrays designed to detect specific rDNA or gene sequences known to be present in important lineages of microorganisms that are thought to be in a specific microbial community (104–106). Loy

and coworkers successfully developed and used a microarray consisting of 132 16S rRNA gene-targeted oligonucleotide probes covering all recognized lineages of sulfate reducing prokaryotes (SRP) for high-resolution screening of clinical and environmental samples (107, 108). The microarray, named SRP- PhyloChip, has great potential for rapid screening of SRP diversity in complex samples and microarray SRP diversity fingerprints allow identification of relevant probes for further characterization of a sample by PCR or quantitative hybridization. This is a valuable option if large numbers of samples are to be analyzed for temporal or spatial variations in SRP diversity. While this approach can provide a comprehensive survey for the presence or absence of a particular sequence, the technique has a closed architecture and cannot identify novel sequences nor can microarrays easily distinguish between two or more closely related sequences in mixed samples.

15.4.2.4 *Serial analysis of ribosomal sequence tags*

Serial analysis of ribosomal sequence tags (SARST) is a novel, powerful approach for high-throughput profiling of microbial diversity in medical, industrial, or environmental samples (1). With SARST, ribosomal sequence tags (RSTs) from hyper-variable V1-region of the bacterial small subunit 16S rRNA gene are amplified using the biotinylated primers Bac64f-BpmI and Bac104r-BsgI. After amplification, the PCR products are digested with BpmI and BsgI, both types of IIS restriction enzymes that will selectively remove the conserved priming regions flanking the V1 region. Subsequently, DNA linkers with compatible ends are ligated to each site of the SARST tags, this to introduce SpeI and NheI restriction sites to flank the SARST tags. Enzymatic digestion with SpeI and NheI is used to create compatible overhangs for SARST tag ligation into concatemers. Subsequently, concatemers are cloned and sequenced, resulting in up to 20 SARST tags per sequencing reaction, hence offering a significant increase in throughput over traditional rDNA clone libraries. The concatemers contain predicted border sequences that demarcate the position and polarity of the SARST tags.

In addition to the protocol by Neufeld and coworkers (109), a simplified method based on the same principles was developed by (110) that uses the V6 variable region to generate SARST tags. SARST libraries were successfully generated from a defined mixture of pure cultures and from duplicate arctic soil DNA samples. The results showed that SARST is functional, reproducible, and that the distribution of RSTs in the defined community library reflects the composition of the original sample. Specific primers could be designed based on sequence data from abundant soil SARST tags for further phylogenetic analysis. However, the distribution of sequences in SARST libraries is biased by multiple copies and similarities between 16S rRNA gene sequences of different species, thereby frustrating accurate quantitative inferences. Nevertheless, the SARST approach is accurate enough to reflect population trends and make comparisons between samples or treatments based on increase or decrease in specific sequence representation.

An alternative for SARST is Single Point Genome Signature Tags (SP-GST), a generally applicable, high-throughput sequencing-based method that targets specific genes to generate identifier tags from well defined points in a genome, was developed (111). The technique yields identifier tags that can distinguish between closely related bacterial strains and allow for the identification of microbial community members. SP-GSTs are determined by three parameters: (i) the primer designed to recognize a conserved gene sequence; (ii) the

anchoring enzyme recognition sequence; (iii) the type IIS restriction enzyme which defines the tag length. The SP-GST method was evaluated *in silico* for bacterial identification using *rpoC*, *uvrB*, *recA* and the 16S rRNA gene. The best distinguishing tags were obtained with the restriction enzyme *Csp6I* upstream of the 16S rRNA gene, which discriminated all organisms in the data set to at least the genus level, and most organisms to the species level. The method was successfully used to generate *Csp6I*-based tags upstream of 16S rRNA gene and allowed to discriminate between closely related strains of *Bacillus cereus* and *Bacillus anthracis*.

15.4.2.5 Paired-end genomic signature tags to study complex microbial communities

The new genomic signature tag (GST)-based DNA tag sequencing and mapping strategies provide a robust tool for comprehensive DNA-based molecular identification and assessment of microorganisms isolated from different habitats. This technology (112) is potentially 30- to 40-fold more efficient than standard DNA sequencing approaches for culture-independent characterization of microbial populations and for detecting genetic rearrangements of model organisms after prolonged periods of growth under different environmental conditions. This methodology can be combined with chromatin immunoprecipitation to identify the *in vivo* chromosomal targets of prokaryotic and eukaryotic transcription factors to aid in elucidating transcriptional regulatory networks underlying global gene expression programs (113).

A new, powerful sequence-based method has been developed to identify and quantitatively analyze genomic DNA or mixtures of genomic DNAs and for mapping the *in vivo* DNA targets for transcription factors. The process based on GST method for amplifying, cloning, and serially analyzing short GST identifier sequences. The improved method uses two recent complementary advances to increase the power of the GST approach. One is the development of new plasmid vectors and cloning strategies that covalently link the 5' and 3' tags from any DNA or cDNA fragment to form what is referred to as a "ditag." The other uses GST tags to map the locations of chromatin immunoprecipitated DNA (ChIP) fragments back to their genome source (113). Combining these procedures allows one for the accurate monitoring of potential changes in rhizosphere genome structure in responses to environmental conditions that may increase mutation rates, thereby, potentially altering rates of evolution *in situ*.

15.4.3 Microarray methods suitable for biomass sampling

DNA microarrays are powerful and versatile tools for monitoring the expression of tens of thousands of genes simultaneously. This technology has successfully been applied to monitor transcriptome regulation in cancer studies (114, 115), the discovery of drug targets (116), and importantly for studying microbial gene expression and regulation under different growth conditions (117). Similar to the situation in which microprocessors have sped up computation, microarray-based genomic technologies have revolutionized the genetic analysis of biological systems, by moving from assaying gene expression, one gene at a time,

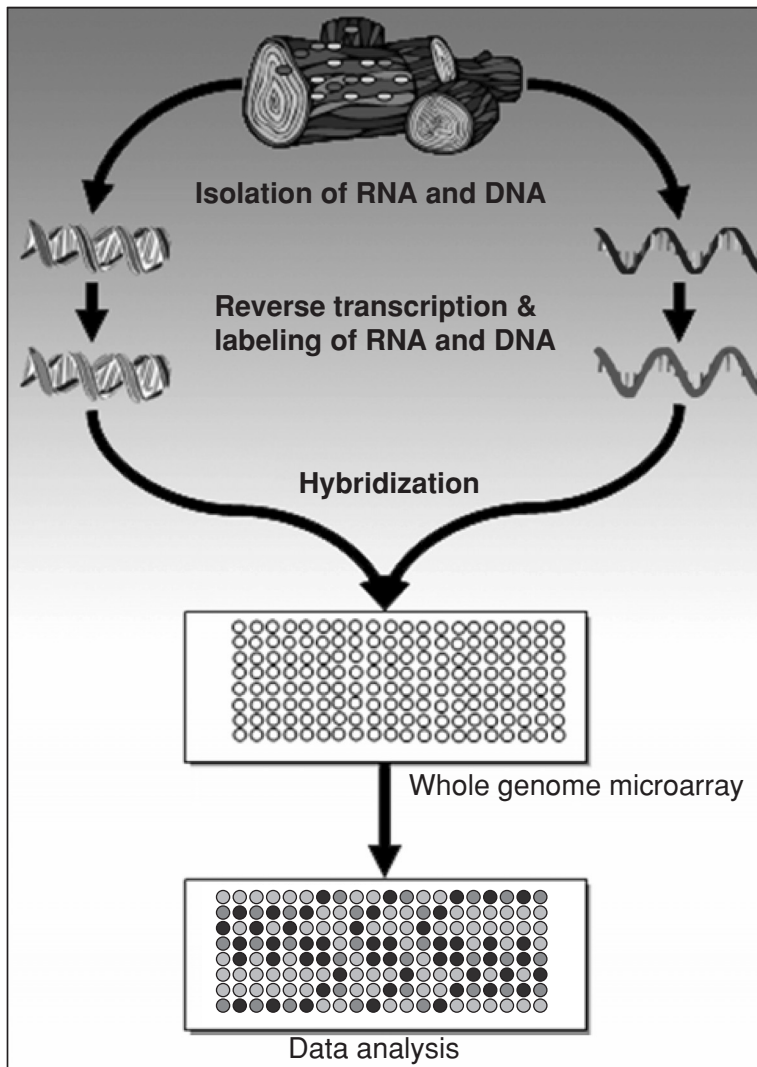


Figure 15.3 Example of a microarray hybridization experiment using a DNA reference. Depending upon the label used an expressed gene can be differentiated from silent genes based on the spot color.

to the ability to visualize the dynamics of the entire transcriptome of an organism in one hybridization step.

The analysis of gene expression using microarrays involves the steps outlined in Figure 15.3. The first step is the acquisition and purification of RNA of samples grown under different experimental conditions. Depending on the experimental design this may also involve the isolation of DNA. Methods for simultaneous isolation and purification of both RNA and DNA (118) from the same sample have been developed. After the isolation of nucleic acids from the sample, a fluorescent label must be incorporated into each. In the case of RNA,

labeled dUTPs are incorporated into the sample, as it is reverse transcribed into cDNA. DNA labeling is accomplished using the Klenow polymerase fragment to incorporate the labeled nucleotide. Following this simultaneous hybridization and subsequent laser excitation and imaging of each of the dye-labeled samples simultaneously, allows one condition to be compared to the other. Transcribed RNA can be compared to baseline DNA from the same sample allowing the gene expression level to be calculated, relative to the initial sample composition.

15.5 Conclusions

The use of lignocellulosic material as a feedstock for producing fuels and chemicals has enormous potential considering its relative abundance. However, its use continues to be cost prohibitive because of pretreatment and enzyme costs (119). What lacks is a fundamental understanding of how natural systems efficiently degrade and utilize decaying biomass. Microorganisms of various types are intimately involved in biomass degradation and occasionally are the only biological agents capable of doing so. Gaining a fundamental understanding of biological mechanisms of biomass decay by microbial communities will help us develop cost-effective ways to convert these abundant substrates to fuels and chemicals.

Microorganisms are not alone in nature. Each microorganism involved in biomass degradation interacts with both its surroundings and with other organisms. These interactions result in chemical and physical changes that in turn lead to microbial communities that are complex and dynamic but ultimately effective in transforming biomass. Conventional microbiological methods used to evaluate them fail to capture their diversity or biochemical complexity. This is due in part because cultivating microorganisms in the laboratory depends on supplying the right nutrients and growth conditions. Because biomass conversion is a dynamic process there are countless microenvironments that would need to be considered.

New methods for examining the diversity and biochemistry of biomass degradation are emerging. It is now possible to characterize microbial populations using large-scale integrative approaches. Rapid advances in the fields of genomics and proteomics have spawned multiple new “omic” subdisciplines as evidenced by the number of recent papers detailing the evolution and molecular basis of microbial communities. Large-scale proteomics-level examination of natural microbial communities is now possible allowing one for not only the functional analysis of enzymes directly involved in biomass conversion but also for the analysis of the processes and interactions involved in community formation. These new technologies will revolutionize the way microbial communities are defined in the future and allow microbiologists to envision and model microbial biomass decay as a set of interacting processes that when combined effectively degrade plant biomass.

Acknowledgment

This work was supported by the US DOE Office of the Biomass Program. The research was also financed as part of the BioEnergy Science Center, a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

References

1. Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F.T., Moore III, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V. & Steffen, W. (2000) The global carbon cycle: A test of our knowledge of earth as a system. *Science*, **290**(5490), 291.
2. Torsvik, V. & Ovreas, L. (June 2002) Microbial diversity and function in soil: From genes to ecosystems. *Current Opinion in Microbiology*, **5**(3), 240–245.
3. Torsvik, V., Ovreas, L. & Thingstad, T.F. (May 10, 2002) Prokaryotic diversity–magnitude, dynamics, and controlling factors. *Science*, **296**(5570), 1064–1066.
4. Raven, J.A. (June 2000) Land plant biochemistry. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*, **355**(1398), 833–846.
5. Henrissat, B. & Romeu, A. (October 1995) Families, superfamilies and subfamilies of glycosyl hydrolases. *Biochemical Journal*, **311**, 350–351.
6. Davies, G.J., Gloster, T.M. & Henrissat, B. (December 2005) Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology*, **15**(6), 637–645.
7. ten Have, R. & Teunissen, P.J. (November 2001) Oxidative mechanisms involved in lignin degradation by white-rot fungi. *Chemical Reviews*, **101**(11), 3397–3413.
8. Perez, J., Munoz-Dorado, J., de la Rubia, T. & Martinez, J. (June 2002) Biodegradation and biological treatments of cellulose, hemicellulose and lignin: An overview. *International Microbiology*, **5**(2), 53–63.
9. Linder, M.B., Szilvay, G.R., Nakari-Setälä, T. & Penttilä, M.E. (2005) Hydrophobins: The protein-amphiphiles of filamentous fungi. *FEMS Microbiology Reviews*, **29**(5): 877–896.
10. Szilvay, G.R., Nakari-Setälä, T. & Linder, M.B. (July 2006) Behavior of *Trichoderma reesei* hydrophobins in solution: Interactions, dynamics, and multimer formation. *Biochemistry*, **45**(28), 8590–8598.
11. Lugones, L.G., Wosten, H.A.B., Birkenkamp, K.U., Sjollem, K.A., Zagers, J. & Wessels, J.G.H. (May 1999) Hydrophobins line air channels in fruiting bodies of *Schizophyllum commune* and *Agaricus bisporus*. *Mycological Research*, **103**, 635–640.
12. Wosten, H.A.B., van Wetter, M.A., Lugones, L.G., van der Mei, H.C., Busscher, H.J. & Wessels, J.G.H. (January 1999) How a fungus escapes the water to grow into the air. *Current Biology*, **9**(2), 85–88.
13. Kostyukovskii, V.A., Okunev, O.N. & Tarakanov, B.V. (November–December 1990) Anaerobic cellulolytic fungi from cattle rumen. *Microbiology*, **59**(6), 749–754.
14. Thormann, M.N., Currah, R.S. & Bayley, S.E. (March 2003) Succession of microfungus assemblages in decomposing peatland plants. *Plant and Soil*, **250**(2), 323–333.
15. Blanchette, R.A., Held, B.W., Jurgens, J.A., McNew, D.L., Harrington, T.C., Duncan, S.M., & Farrell, R.L. (March 2004) Wood-destroying soft rot fungi in the historic expedition huts of Antarctica. *Applied and Environmental Microbiology*, **70**(3), 1328–1335.
16. Blanchette, R.A., Krueger, E.W., Haight, J.E., Akhtar, M. & Akin, D.E. (March 14, 1997) Cell wall alterations in loblolly pine wood decayed by the white-rot fungus, *Ceriporiopsis subvermispora*. *Journal of Biotechnology*, **53**(2–3), 203–213.
17. Redlak, K., Dahm, H., Ciesielska, A. & Strzelczyk, E. (2001) Enzymatic activity of ectendomycorrhizal fungi. *Biology and Fertility of Soils*, **33**(1), 83–90.
18. Hohnjec, N., Vieweg, M.E., Puhler, A., Becker, A. & Kuster, H. (April 2005) Overlaps in the transcriptional profiles of *Medicago truncatula* roots inoculated with two different *Glomus* fungi provide insights into the genetic program activated during arbuscular mycorrhiza. *Plant Physiology*, **137**(4), 1283–1301.

19. Garcia-Garrido, J.M., Tribak, M., Rejon-Palomares, A., Ocampo, J.A. & Garcia-Romera, I. (August 2000) Hydrolytic enzymes and ability of arbuscular mycorrhizal fungi to colonize roots. *Journal of Experimental Botany*, **51**(349), 1443–1448.
20. Coulthurst, S.J., Kurz, C.L. & Salmond, G.P.C. (2004) luxS mutants of *Serratia* defective in autoinducer-2-dependent “quorum sensing” show strain-dependent impacts on virulence and production of carbapenem and prodigiosin. *Microbiology-SGM*, **150**, 1901–1910.
21. Burrell, P.C. (2006) The detection of environmental autoinducing peptide quorum-sensing genes from an uncultured *Clostridium* sp. in landfill leachate reactor biomass. *Letters in Applied Microbiology*, **43**(4), 455–460.
22. Reverchon, S., Bouillant, M.L., Salmond, G. & Nasser, W. (1998) Integration of the quorum-sensing system in the regulatory networks controlling virulence factor synthesis in *Erwinia chrysanthemi*. *Bioorganic & Medicinal Chemistry Letters*, **12**(8), 1407–1418.
23. Gonzalez, J.E. & Marketon, M.M. (2003) Quorum sensing in nitrogen-fixing Rhizobia. 574–592.
24. Beguin, P. (1990) Molecular biology of cellulose degradation. *Annual Review of Microbiology*, **44**, 219–248.
25. Beguin, P., Millet, J. & Aubert, J.P. (1992) Cellulose degradation by *Clostridium thermocellum*: From manure to molecular biology. *FEMS Microbiology Letters*, **100**(1–3), 523–528.
26. Kluepfel, D. (1988) Screening of prokaryotes for cellulose and hemicellulose degrading enzymes. In: *Methods in Enzymology: Biomass Part A Cellulose and Hemicellulose* (eds S.T. Kellogg & W.A. Wood), p. 180. Academic Press, San Diego, CA.
27. Pankratov, T.A., Dedys, S.N. & Zavarzin, G.A. (2006) The leading role of actinobacteria in aerobic cellulose degradation in Sphagnum peat bogs. *Doklady Biological Sciences*, **410**(1), 428–430.
28. Mohagheghi, A., Grohmann, K., Himmel, M., Leighton, L. & Updegraff, D.M. (July 1986) Isolation and characterization of acidothermus-cellulolyticus Gen-Nov, Sp-Nov, a new genus of thermophilic, acidophilic, cellulolytic bacteria. *International Journal of Systematic Bacteriology*, **36**(3), 435–443.
29. Ding, S.Y., Adney, W.S., Vinzant, T.B., Decker, S.R., Baker, J.O., Thomas, S.R., & Himmel, M.E. (2003) Glycoside hydrolase gene cluster of *Acidothermus cellulolyticus*. *Applications of Enzymes to Lignocellulosics*, 332–360.
30. Schwintzer, C.R. & Tjepkema, J.D. (1991) The biology of Frankia and Actinorhizal plants. *The Quarterly Review of Biology*, **66**(1), 84–85.
31. Benson, D.R. & Silvester, W.B. (June 1993) Biology of Frankia strains, actinomycete symbionts of actinorhizal plants. *Microbiological Reviews*, **57**(2), 293–319.
32. Hugenholtz, P., Goebel, B.M. & Pace, N.R. (September 15, 1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology*, **180**(18), 4765–4774.
33. Amann, R.L., Ludwig, W. & Schleifer, K.H. (March 1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiology Reviews*, **59**(1), 143–169.
34. Ward, D.M., Weller, R. & Bateson, M.M. (May 1990) 16S ribosomal-RNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature*, **345**(6270), 63–65.
35. Torsvik, V., Goksoyr, J. & Daae, F.L. (March 1, 1990) High diversity in DNA of soil bacteria. *Applied and Environmental Microbiology*, **56**(3), 782–787.
36. Torsvik, V., Sorheim, R. & Goksoyr, J. (September–October 1996) Total bacterial diversity in soil and sediment communities – A review. *Journal of Industrial Microbiology*, **17**(3–4), 170–178.
37. Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. & Goodman, R.M. (October 1998) Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chemistry and Biology*, **5**(10), R245–R249.
38. Chang, W.T.H. & Thayer, D.W. (1977) Cellulase system of a cytophaga species. *Canadian Journal of Microbiology*, **23**(9), 1285–1292.

39. McBride, M.J. & Baker, S.A. (August 1996) Development of techniques to genetically manipulate members of the genera *Cytophaga*, *Flavobacterium*, *Flexibacter*, and *Sporocytophaga*. *Applied and Environmental Microbiology*, **62**(8), 3017–3022.
40. McBride, M.J. (2001) Bacterial gliding motility: Multiple mechanisms for cell movement over surfaces. *Annual Review of Microbiology*, **55**, 49–75.
41. Pate, J.L. (1985) Gliding motility in cytophaga. *Microbiological Sciences*, **2**(10), 289–291.
42. Taylor, L.E., Henrissat, B., Coutinho, P.M., Ekborg, N.A., Hutcheson, S.W. & Weiner, R.A. (June 2006) Complete cellulase system in the marine bacterium *Saccharophagus degradans* strain 2-40(T). *Journal of Bacteriology*, **188**(11), 3849–3861.
43. Ekborg, N.A., Taylor, L.E., Longmire, A.G., Henrissat, B., Weiner, R.M., & Hutcheson, S.W. (May 2006) Genomic and proteomic analyses of the agarolytic system expressed by *Saccharophagus degradans* 2–40. *Applied and Environmental Microbiology*, **72**(5), 3396–3405.
44. Warnick, T.A., Methe, B.A. & Leschine, S.B. (July 2002) *Clostridium phytofermentans* sp nov., a cellulolytic mesophile from forest soil. *International Journal of Systematic and Evolutionary Microbiology*, **52**, 1155–1160.
45. Grayston, S.J. & Campbell, C.D. (November–December 1996) Functional biodiversity of microbial communities in the rhizospheres of hybrid larch (*Larix eurolepis*) and Sitka spruce (*Picea sitchensis*). *Tree Physiology*, **16**(11–12), 1031–1038.
46. Lynch, J.M. & Whipps, J.M. (December 1990) Substrate flow in the rhizosphere. *Plant and Soil*, **129**(1), 1–10.
47. Garbaye, J. (October 1994) Helper bacteria – a new dimension to the mycorrhizal symbiosis. *New Phytologist*, **128**(2), 197–210.
48. Tien, T.M., Gaskins, M.H. & Hubbell, D.H. (1979) Plant-growth substances produced by azospirillum-brasilense and their effect on the growth of pearl-millet (*Pennisetum-Americanum* L.). *Applied and Environmental Microbiology*, **37**(5), 1016–1024.
49. Gutierrez-Manero, F.J., Ramos-Solano, B., Probanza, A., Mehrouachi, J., Tadeo, F.R. & Talon, M. (February 2001) The plant-growth-promoting rhizobacteria *Bacillus pumilus* and *Bacillus licheniformis* produce high amounts of physiologically active gibberellins. *Physiologia Plantarum*, **111**(2), 206–211.
50. Stroeve, P., Pettit, C.D., Vasquez, V., Kim, I. & Berry, A.M. (July 1998) Surface active behavior of hopanoid lipids: Bacteriohopanetetrol and phenylacetate monoester bacteriohopanetetrol. *Langmuir*, **14**(15), 4261–4265.
51. Smolander, A., Ronkko, R., Nurmiaholaassila, E.L. & Haahtela, K. (September 1990) Growth of frankia in the rhizosphere of betula-pendula, a nonhost tree species. *Canadian Journal of Microbiology*, **36**(9), 649–656.
52. Mansour, S.R. (1994) Production of growth-hormones in casuarina-cunninghamiana root-nodules induced by Frankia strain Hfpcgi4. *Protoplasma*, **183**(1–4), 126–130.
53. Haahtela, K., Kilpi, S. & Kari, K. (May–June 1988) Effects of phenoxy acid herbicides and glyphosate on nitrogenase activity (acetylene-reduction) in root-associated Azospirillum, Enterobacter and Klebsiella. *FEMS Microbiology Ecology*, **53**(3–4), 123–127.
54. Barbieri, E., Potenza, L., Rossi, I., Sisti, D., Giomaro, G., Rossetti, S., & Stocchi, V. (November 2000) Phylogenetic characterization and in situ detection of a Cytophaga-Flexibacter-Bacteroides phylogroup bacterium in Tuber borchii Vittad, ectomycorrhizal mycelium. *Applied and Environmental Microbiology*, **66**(11), 5035–5042.
55. Sarand, I., Timonen, S., Nurmiaho-Lassila, E.L., Koivula, T., Haahtela, K., Romantschuk, M., & Sen, R. (October 1998) Microbial biofilms and catabolic plasmid harbouring degradative fluorescent pseudomonads in Scots pine mycorrhizospheres developed on petroleum contaminated soil. *FEMS Microbiology Ecology*, **27**(2), 115–126.
56. Nurmiaho-Lassila, E.L., Timonen, S., Haahtela, K. & Sen, R. (November 1997) Bacterial colonization patterns of intact Pinus sylvestris mycorrhizospheres in dry pine forest soil: An electron microscopy study. *Canadian Journal of Microbiology*, **43**(11), 1017–1035.

57. Perotto, S., Coisson, J.D., Perugini, I., Cometti, V. & Bonfante, P. (January 1997) Production of pectin-degrading enzymes by ericoid mycorrhizal fungi. *New Phytologist*, **135**(1), 151–162.
58. Garbaye, J. (1994) Mycorrhization helper bacteria – a new dimension to the mycorrhizal symbiosis. *Acta Botanica Gallica*, **141**(4), 517–521.
59. Frey, P., FreyKlett, P., Garbaye, J., Berge, O. & Heulin, T. (May 1997) Metabolic and genotypic fingerprinting of fluorescent pseudomonads associated with the Douglas fir *Laccaria bicolor* mycorrhizosphere. *Applied and Environmental Microbiology*, **63**(5), 1852–1860.
60. Atkinson, C.F., Jones, D.D. & Gauthier, J.J. (Fall 1996) Biodegradabilities and microbial activities during composting of municipal solid waste in bench-scale reactors. *Compost Science and Utilization*, **4**(4), 14–23.
61. Battin, T.J., Kaplan, L.A., Newbold, J.D. & Hansen, C.M.E. (November 27, 2003) Contributions of microbial biofilms to ecosystem processes in stream mesocosms. *Nature*, **426**(6965), 439–442.
62. Maximov, V.V. (September 2006) Microbiological specificity of Baikal and Khubsugul – ancient lakes of Eastern Asia. *Hydrobiologia*, **568**, 63–68.
63. Cullings, K. & Makhija, S. (December 2001) Ectomycorrhizal fungal associates of *Pinus contorta* in soils associated with a hot spring in Norris Geyser basin, Yellowstone National Park, Wyoming. *Applied and Environmental Microbiology*, **67**(12), 5538–5543.
64. Alekhina, L.K., Polyanskaya, L.M. & Dobrovol'skaya, T.G. (January 2001) Population dynamics of microorganisms in the soils of the Central Forest Reserve (model experiments). *Eurasian Soil Science*, **34**(1), 88–91.
65. Atkinson, C.F., Jones, D.D. & Gauthier, J.J. (September 1997) Microbial activities during composting of pulp and paper-mill primary solids. *World Journal of Microbiology and Biotechnology*, **13**(5), 519–525.
66. Buyer, J.S., Roberts, D.P., Millner, P. & Russek-Cohen, E. (May 2001) Analysis of fungal communities by sole carbon source utilization profiles. *Journal of Microbiological Methods*, **45**(1), 53–60.
67. Andersson, B.E., Tornberg, K., Henrysson, T. & Olsson, S. (October 2001) Three-dimensional outgrowth of a wood-rotting fungus added to a contaminated soil from a former gasworks site. *Bioresource Technology*, **78**(1), 37–45.
68. Cahyani, V.R., Watanabe, A., Matsuya, K., Asakawa, S. & Kimura, M. (October 2002) Succession of microbiota estimated by phospholipid fatty acid analysis and changes in organic constituents during the composting process of rice straw. *Soil Science and Plant Nutrition*, **48**(5), 735–743.
69. Andersson, M., Kjoller, A. & Struwe, S. (2004) Microbial enzyme activities in leaf litter, humus and mineral soil layers of European forests. *Soil Biology and Biochemistry*, **36**(10), 1527–1537.
70. Libmond, S. & Savoie, J.M. (December 1993) Degradation of wheat-straw by a microbial community – stimulation by a polysaccharidase complex. *Applied Microbiology and Biotechnology*, **40**(4), 567–574.
71. Graves, P.R. & Haystead, T.A.J. (March 2002) Molecular biologist's guide to proteomics. *Microbiology and Molecular Biology Reviews*, **66**(1), 39–63.
72. Stemmer, M. (2004) Multiple-substrate enzyme assays: A useful approach for profiling enzyme activity in soils? *Soil Biology and Biochemistry*, **36**(3), 519–527.
73. Waldrop, M.P. & Zak, D.R. (September 2006) Response of oxidative enzyme activities to nitrogen deposition affects soil concentrations of dissolved organic carbon. *Ecosystems*, **9**(6), 921–933.
74. Renella, G., Mench, M., van der Lelie, D., Pietramellara, G., Ascher, J., Ceccherini, M.T., Landi, L. & Nannipieri, P. (March 2004) Hydrolase activity, microbial biomass and community structure in long-term Cd-contaminated soils. *Soil Biology and Biochemistry*, **36**(3), 443–451.
75. Allison, S.D. & Jastrow, J.D. (November 2006) Activities of extracellular enzymes in physically isolated fractions of restored grassland soils. *Soil Biology and Biochemistry*, **38**(11), 3245–3256.

76. Voget, S., Leggewie, C., Uesbeck, A., Raasch, C., Jaeger, K.E. & Streit, W.R. (October 1, 2003) Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology*, **69**(10), 6235–6242.
77. Chander, K., Goyal, S., Nandal, D.P. & Kapoor, K.K. (June 1998) Soil organic matter, microbial biomass and enzyme activities in a tropical agroforestry system. *Biology and Fertility of Soils*, **27**(2), 168–172.
78. Ofarrell, P.H. (1975) High-resolution 2-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, **250**(10), 4007–4021.
79. Unlu, M., Morgan, M.E. & Minden, J.S. (October 1997) Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis*, **18**(11), 2071–2077.
80. Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I. & Davison, M. (March 2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*, **1**(3), 377–396.
81. Karp, N.A., Kreil, D.P. & Lilley, K.S. (May 2004) Determining a significant change in protein expression with DeCyder (TM) during a pair-wise comparison using two-dimensional difference gel electrophoresis. *Proteomics*, **4**(5), 1421–1432.
82. Alban, A., David, S.O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S. & Currie, I. (January 2003) A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics*, **3**(1), 36–44.
83. Torsvik, V., Daae, F.L., Sandaa, R.A. & Ovreas, L. (September 1998) Novel techniques for analysing microbial diversity in natural and perturbed environments. *Journal of Biotechnology*, **64**(1), 53–62.
84. Teeri, T.T., Lehtovaara, P., Kauppinen, S., Salovuori, I. & Knowles, J. (1987) Homologous domains in *Trichoderma-Reesei* cellulolytic enzymes – gene sequence and expression of cellobiohydrolase-II. *Gene*, **51**(1), 43–52.
85. Chander, K., Goyal, S. & Kapoor, K.K. (March 1995) Microbial biomass dynamics during the decomposition of leaf-litter of poplar and eucalyptus in a sandy loam. *Biology and Fertility of Soils*, **19**(4), 357–362.
86. Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Hatada, I., Shibata, H., Kawai, J., Hirose, K., Watanabe, S., Fushiki, S., Wada, S., Sugimoto, T., Kobayakawa, K., Kawara, T., Katsuki, M., Shibuya, T. & Mukai, T. (April 1993) Restriction landmark genomic scanning method and its various applications. *Electrophoresis*, **14**(4), 251–258.
87. Fisher, M.M. & Triplett, E.W. (October 1999) Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Applied and Environmental Microbiology*, **65**(10), 4630–4636.
88. Lesaulnier, C., Papamichail, D., McCorkle, S., Ollivier, B., Skiena, S., Taghavi, S., Zak, D. & van der Lelie, D., (2008) Elevated CO₂ affects soil microbial diversity associated with trembling aspen. *Environmental Microbiology*, in press.
89. Muyzer, G., de Waal, E.C. & Uitterlinden, A.G. (March 1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, **59**(3), 695–700.
90. Ranjard, L., Poly, F., Combrisson, J., Richaume, A., Gourbiere, F., Thioulouse, J. & Nazaret, S. (May 2000) Heterogeneous cell density and genetic structure of bacterial pools associated with various soil microenvironments as determined by enumeration and DNA fingerprinting approach (RISA). *Microbial Ecology*, **39**(4), 263–272.
91. Ranjard, L., Poly, F. & Nazaret, S. (April 2000) Monitoring complex bacterial communities using culture-independent molecular techniques: Application to soil environment. *Research in Microbiology*, **151**(3), 167–177.

92. Ranjard, L., Echairi, A., Nowak, V., Lejon, D.P.H., Nouaim, R. & Chaussod, R. (November 2006) Field and microcosm experiments to evaluate the effects of agricultural Cu treatment on the density and genetic structure of microbial communities in two different soils. *FEMS Microbiology Ecology*, **58**(2), 303–315.
93. Ranjard, L., Brothier, E. & Nazaret, S. (December 2000) Sequencing bands of ribosomal intergenic spacer analysis fingerprints for characterization and microscale distribution of soil bacterium populations responding to mercury spiking. *Applied and Environmental Microbiology*, **66**(12), 5334–5339.
94. Ovreas, L., Jensen, S., Daae, F.L. & Torsvik, V. (July 1998) Microbial community changes in a perturbed agricultural soil investigated by molecular and physiological approaches. *Applied and Environmental Microbiology*, **64**(7), 2739–2742.
95. Tringe, S.G. & Rubin, E.M. (November 2005) Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics*, **6**(11), 805–814.
96. Schloss, P.D. & Handelsman, J. (2003) Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology*, **14**(3), 303–310.
97. Breitbart, M., Felts, B., Kelley, S., Mahaffy, J.M., Nulton, J., Salamon, P. & Rohwer, F. (March 2004) Diversity and population structure of a near-shore marine-sediment viral community. *Proceedings of the Royal Society of London Series B – Biological Sciences*, **271**(1539), 565–574.
98. Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P. & Rohwer, F. (October 2003) Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology*, **185**(20), 6220–6223.
99. Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F. & Rohwer, F. (October 2002) Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(22), 14250–14255.
100. Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D. S. & Banfield, J.F. (March 2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, **428**(6978), 37–43.
101. Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. & Smith, H.O. (April 2, 2004) Environmental genome shotgun sequencing of the sargasso sea. *Science*, **304**(5667), 66–74.
102. Curtis, T.P. & Sloan, W.T. (June 2004) Prokaryotic diversity and its limits: Microbial community structure in nature and implications for microbial ecology. *Current Opinion in Microbiology*, **7**(3), 221–226.
103. Curtis, T.P., Sloan, W.T. & Scannell, J.W. (August 2002) Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(16), 10494–10499.
104. Schena, M., Heller, R.A., Thieriault, T.P., Konrad, K., Lachenmeier, E. & Davis, R.W. (July 1998) Microarrays: Biotechnology's discovery platform for functional genomics. *Trends in Biotechnology*, **16**(7), 301–306.
105. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. (October 1995) Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. *Science*, **270**(5235), 467–470.
106. Small, J., Call, D.R., Brockman, F.J., Straub, T.M. & Chandler, D.P. (2001) Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. 4708–4716.
107. Loy, A., Kusel, K., Lehner, A., Drake, H.L. & Wagner, M. (December 2004) Microarray and functional gene analyses of sulfate-reducing prokaryotes in low-sulfate, acidic fens reveal cooccurrence of recognized genera and novel lineages. *Applied and Environmental Microbiology*, **70**(12), 6998–7009.

108. Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J., Schleifer, K.H. & Wagner, M. (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Applied and Environmental Microbiology*, **68**, 5064–5081.
109. Neufeld, J.D., Yu, Z., Lam, W. & Mohn, W.W. (February 2004) Serial analysis of ribosomal sequence tags (SARST): A high-throughput method for profiling complex microbial communities. *Environmental Microbiology*, **6**(2), 131–144.
110. Kysela, D.T., Palacios, C. & Sogin, M.L. (2005) Serial analysis of V6 ribosomal sequence tags (SARST-V6): A method for efficient, high-throughput analysis of microbial community composition. *Environmental Microbiology*, **7**(3), 356–364.
111. van der Lelie, D., Lesaulnier, C., McCorkle, S., Geets, J., Taghavi, S. & Dunn, J. (2006). Use of single-point genome signature tags as a universal tagging method for microbial genome surveys. *Applied Environmental Microbiology*, **72**, 2092–2101.
112. Dunn, J.J., McCorkle, S.R., Praissman, L.A., Hind, G., van der Lelie, D., Bahou, W.F., Gnatenko, D.V. & Krause, M.K. (November 1, 2002) Genomic Signature Tags (GSTs): A system for profiling genomic DNA. *Genome Research*, **12**(11), 1756–1765.
113. Impey, S., McCorkle, S.R., Cha-Molstad, H., Dwyer, J.M., Yochum, G.S., Boss, J.M., McWeeney, S., Dunn, J.J., Mandel, G. & Goodman, R.H. (December 2004) Defining the CREB regulon: A genome-wide analysis of transcription factor regulatory regions. *Cell*, **119**(7), 1041–1054.
114. Liotta, L. & Petricoin, E. (October 2000) Molecular profiling of human cancer. *Nature Reviews Genetics*, **1**(1), 48–56.
115. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, R., Caligiuri, M.A., Bloomfield, C.D. & Lander, E.S. (October 1999) Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science*, **286**(5439), 531–537.
116. Debouck, C. & Goodfellow, P.N. (January 1999) DNA microarrays in drug discovery and development. *Nature Genetics*, **21**, 48–50.
117. Conway, T. & Schoolnik, G.K. (February 2003) Microarray expression profiling: Capturing a genome-wide portrait of the transcriptome. *Molecular Microbiology*, **47**(4), 879–889.
118. Hurt, R.A., Qiu, X.Y., Wu, L.Y., Roh, Y., Palumbo, A.V., Tiedje, J.M. & Zhou, J.H. (October 2001) Simultaneous recovery of RNA and DNA from soils and sediments. *Applied and Environmental Microbiology*, **67**(10), 4495–4503.
119. Sheehan, J. & Himmel, M. (September–October 1999) Enzymes, energy, and the environment: A strategic perspective on the US Department of Energy's Research and Development Activities for Bioethanol. *Biotechnology Progress*, **15**(5), 817–827.

Chapter 16

New Generation Biomass Conversion: Consolidated Bioprocessing

Y.-H. Percival Zhang and Lee R. Lynd

16.1 Introduction

Accumulation of the greenhouse gas, CO₂, mainly from burning of fossil fuels, and the depletion of finite fossil fuels are vital threats to the sustainable development of humans (1–3). Lignocellulose is the most abundant renewable biological resource today (ca. 2×10^{11} tons/year) and is produced by photosynthesis (i.e., plants fix atmospheric CO₂) (4–8). Development of technologies for effectively converting less-costly agricultural and forestry residues for use in bio-based chemical and fuels production offers potential benefits to the national interest by improving strategic security, decreasing trade deficits, encouraging healthier rural economies, and improving environmental quality by moving closer to zero net greenhouse gas emissions and sustainable resource supplies (1–3, 9–19).

Lignocellulosic feedstock is far less costly than other energy feedstocks (i.e., crude oil, natural gas, corn kernels, and soy oils) based on energy content (\$/GJ) in Figure 16.1. For example, when crude oil prices vary from \$40 to \$70 per barrel, equaling \$7.1 to \$12.1/GJ, they are much higher than those of lignocellulose (\$0 to \$3/GJ). Similarly, corn kernels costing from \$2.25 to \$4.0 per bushel equal \$6.3 to \$11.5/GJ. During the past 2 years (2004–2006), corn kernel prices have risen by >70% from the historically low prices (~\$2.25 per bushel) to ~\$4 per bushel. With high demands of corn kernels for ethanol production, higher prices of corn kernels have resulted in rising prices of animal feed and human food. For example, the Chinese government banned the establishment of new ethanol production facilities based on grains in 2006. As expected, less costly feedstock and the most abundant supplies make production of fuels and chemicals from lignocellulose appealing.

Production of commodity products (i.e., fuels, chemicals, and materials) from renewable biomass is distinct from biotechnology motivated by health care at many levels, including economic driving forces, importance of feedstock prices, processing costs and capital investment, the scale of applications, and feedstock availability (10). Biocommodities have low selling prices so that raw material costs are often dominant factors in determining prices of commodity products (~30–70%), whereas raw materials usually account for a very small fraction of high selling prices of pharmaceuticals (10, 13). The production costs, including capital recovery and processing costs, are usually another dominant factor determining the

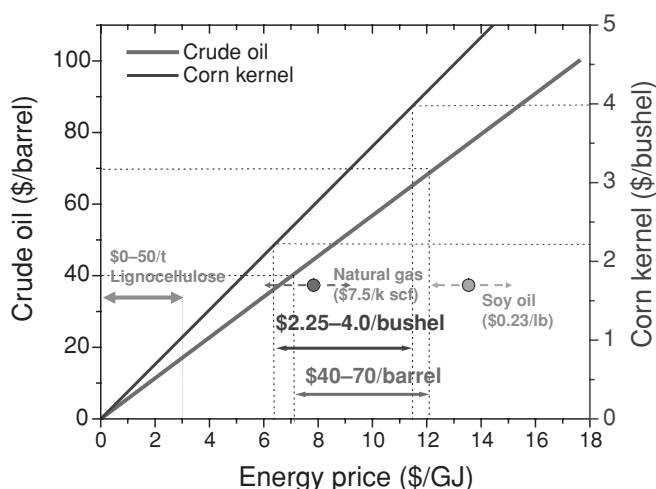


Figure 16.1 Comparison of prices of the feedstocks for liquid transportation fuels.

price of commodity products, whereas they are not nearly so important for pharmaceuticals. Market sizes for individual biocommodity products and biopharmaceuticals are of relatively similar magnitude. However, tremendous differences exist with respect to a product mass basis, e.g., the largest commodity markets exceed pharmaceutical markets by approximately 11 orders of magnitude (10). The production of high-volume/low-value biocommodity products has an absolute requirement for high-volume/low-cost feedstock and must be responsive to the availability and characteristics of feedstock, whereas no such requirement exists for the production of pharmaceuticals.

Overcoming the recalcitrant structure of lignocellulose is still among the greatest challenges for the emerging biofuel and bio-based chemical industries (20, 21). Currently, high conversion costs, large investment risks, and a narrow economic margin between feedstock costs and product prices slow the realization of cellulosic ethanol production on a large scale (22–24). Effective biological conversion of recalcitrant lignocellulose to biocommodity products involves four main sequential steps: 1) biomass size reduction, 2) pretreatment/fractionation, 3) enzymatic cellulose and hemicellulose hydrolysis, and 4) fermentation (11, 24, 25). For most types of lignocellulose biomass, the enzymatic digestibility of cellulose is very low (<20%) without some type of pretreatment that opens up the structure and makes it accessible to attack by enzymes (11, 24, 25). A number of biological, chemical, and physical pretreatment techniques have been investigated (22, 23, 26, 27).

16.2 Consolidated bioprocessing

After lignocellulose pretreatment, there are four biologically mediated events typically in the course of biological processing of cellulosic biomass: cellulase production, enzymatic cellulose hydrolysis, hexose fermentation, and pentose fermentation (Figure 16.2). Separate

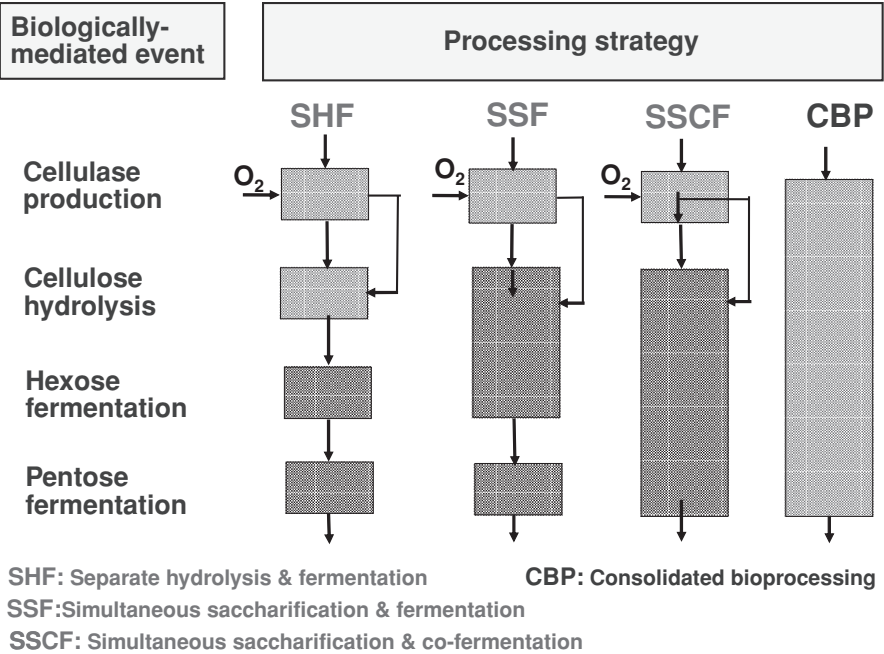


Figure 16.2 Evolution of biomass processing configurations featuring enzymatic hydrolysis.

hydrolysis and fermentation (SHF) involves these four discrete process steps. Simultaneous saccharification and fermentation (SSF) consolidates cellulose hydrolysis and hexose fermentation. Simultaneous saccharification and co-fermentation (SSCF) combines cellulose hydrolysis, hexose fermentation, and pentose fermentation. Consolidated bioprocessing (CBP) integrates cellulase production, cellulose hydrolysis, with pentose and hexose fermentations in a single step (10, 15, 28).

Over the past few years, much effort has been devoted to reducing the cost of cellulase enzyme production (24). Following greater than 20-fold cost reductions, cellulase production costs have recently been reported in the range of ~20 cents per gallon of cellulosic ethanol produced (29). These developments enable a variety of formerly infeasible industrial SSF and SSCF processes, but do not diminish the competitive potential of CBP by offering significantly lower costs than other processes.

CBP offers the potential for lower production costs, lower capital investment, and higher conversion efficiency as compared to the processes featuring dedicated cellulase production. CBP avoids costs for capital, substrate, other raw materials, and utilities associated with cellulase production. In addition, CBP could realize higher hydrolysis rates, and hence reduce reactor volume and capital investment as a result of enzyme–microbe synergy. CBP provides access to the use of thermophiles or other organisms with high activity cellulases. Moreover, cellulose-adherent cellulolytic microorganisms may successfully compete for products of cellulose hydrolysis with non-adhered microbes. Moreover, these microorganisms are likely to be less sensitive to contaminants, which could increase the stability of an industrial processes. Economic analysis suggests that the sum of 9.9 ¢/gal ethanol

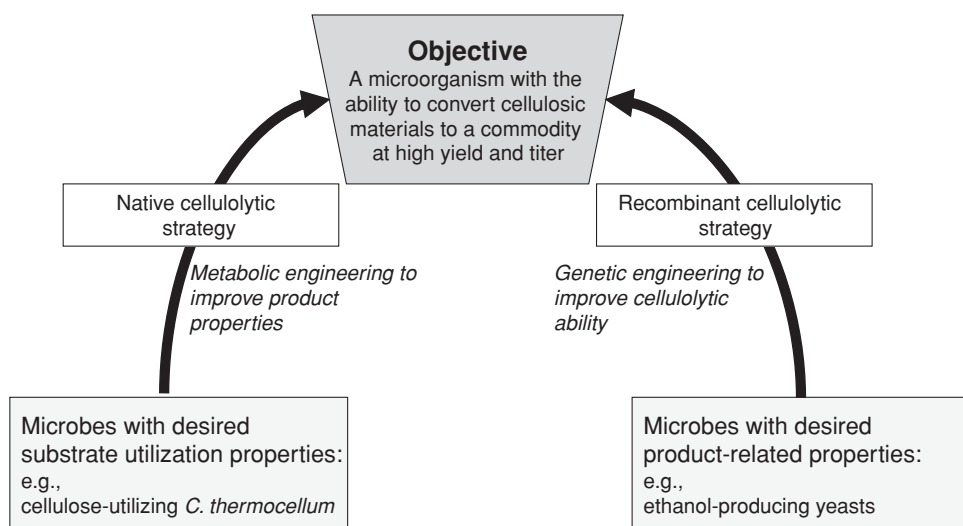


Figure 16.3 Alternative organism development strategies to obtain organisms useful in processing cellulosic feedstocks.

for dedicated cellulase production and 9.0 ¢/gal for SSCF gives a total cost for biological processing of 18.9 ¢/gal, which is more than fourfold greater than the 4.2 ¢/gal projected for CBP (28).

Today, there are no CBP-enabling microorganisms suitable for industrial applications. CBP microorganism development can proceed via a native cellulose utilization strategy and a recombinant cellulose utilization strategy (Figure 16.3). The native cellulolytic strategy involves engineering product metabolism to produce desired products based on naturally cellulolytic microorganisms (e.g., *Clostridium thermocellum*). The recombinant cellulolytic strategy involves introducing heterologous cellulase genes into an organism whose product yield and tolerance credentials are well-established (e.g., Baker's yeast *Saccharomyces cerevisiae*). Each strategy has its own advantages and challenges, and different strategies may well prove most advantageous for different products.

16.3 CBP advances

16.3.1 Native cellulolytic microorganisms

C. thermocellum, an anaerobic, thermophilic, Gram-positive bacterium, exhibits one of the highest rates of cellulose utilization among described microorganisms (15). *C. thermocellum* produces a cellulase complex, or "cellulosome," a substantial fraction of which is bound to the cell surface under most culture conditions (30–33). Because the turnover number of cellulases on insoluble substrates is much lower than most catalytic enzymes on soluble substrates, substantial amounts of cellulase (~2–20% of cellular proteins, weight based) are required to support cellulolytic cells growing on cellulose. Because cellulose is an insoluble

substrate, cellulases must be secreted across the cell membrane so that they can access and hydrolyze insoluble cellulose to soluble sugars that cells can assimilate.

16.3.1.1 Cellulase regulation

Regulation of cellulase synthesis by *C. thermocellum* is an important feature of the physiology of this microorganism, particularly in light of the substantial investment of ATP that cellulase synthesis represents (34, 35). Johnson and coworkers (36) reported that true cellulase activity (i.e., degradation of crystalline cellulose) synthesis was markedly repressed by cellobiose. mRNAs corresponding to endoglucanases CelA, CelF, and CelD were found to be regulated at the level of transcription by a mechanism analogous to catabolite repression (37). The number of CelS (the most dominant cellulolytic enzyme of the cellulosome) and CipA (the cellulosomal scaffolding protein) transcripts per cell were shown to decrease with increasing growth rate (38, 39). CelS transcripts were found to be higher for growth under cellobiose-limitation as compared to growth under nitrogen limitation (38) and control of scaffoldin and CelS transcription was shown to involve a housekeeping Sigma-A factor (39).

Stevenson and Weimer (40) investigated expression profiles of 17 genes involved in cellulose hydrolysis, intracellular phosphorylation, catabolite repression, and fermentation end product formation as determined by real-time PCR in continuous cultures grown on cellobiose and cellulose. Thirteen genes displayed modest (fivefold or less) differences in expression in response to varied growth rate or substrate. By contrast, *cipA*, *celS*, and *manA* genes displayed 10-fold reduced levels when grown on cellobiose at dilution rates of $>0.05/\text{h}$, suggesting that at least some cellulosomal components are transcriptionally regulated.

Zhang and Lynd (32) investigated the regulation of cell-specific cellulase synthesis (defined as mg cellulase/g cell dry weight) by *C. thermocellum* using an ELISA protocol based on an antibody raised against a peptide sequence from the scaffoldin protein (31). We found that cellulase synthesis in Avicel-grown batch cultures was nine times greater than in cellobiose-grown batch cultures. In substrate-limited continuous cultures, however, cellulase synthesis with Avicel-grown cultures was greater than that in cellobiose-grown cultures by 1.3- to 2.4-fold, depending on the dilution rate. Continuous cellobiose-grown cultures maintained at either high dilution rates or high feed substrate concentration resulted in decreased cellulase synthesis, with a large (sevenfold) decrease between 0 and 0.2 g/L cellobiose and a much more gradual further decrease for cellobiose concentrations >0.2 g/L. Several factors suggest that cellulase synthesis in *C. thermocellum* is regulated by carbon catabolite repression (CCR). These factors include: 1) substantially higher cellulase yields observed during batch growth on Avicel as compared to cellobiose, 2) a strong negative correlation between cellobiose concentration and cellulase yield in continuous cultures with varied dilution rate at constant feed substrate concentration and also with varied feed substrate concentration at constant dilution rate, and 3) the presence of sequences corresponding to key elements of catabolite repression systems in the *C. thermocellum* genome. CCR-mediated control of cellulosome synthesis in *C. thermocellum* is supported by the observation that the three key components of a CCR system – an LacI/GalR family regulatory protein, an HPr protein and an HPr kinase, and a 14-bp *cis*-acting catabolite responsive element (CRE) binding sequence – are present in the *C. thermocellum* genomic sequence (<http://genome.ornl.gov/microbial/cthe/>). Several putative LacI/GalR family genes are found in *C. thermocellum*. We were able to locate many (>100) putative CRE sequences, including two putative CREs inside the *cipA*

structural gene (+953, +5231), using the more degenerate CRE consensus sequence (WG-WNANC/GNTNNCW). Substantial degeneracy of CRE sequences is supported by results from *B. subtilis*. For example, Chaveaux (41) found only 29 CRE sequences based on a consensus sequence with 7 of the 14 bases degenerate, whereas Moreno and coworkers (42) found, using DNA arrays, that ~330 genes were regulated by CCR. Moreover, whole genome analysis of *B. subtilis* indicates that the CREs sequence is not strictly conserved, and that CRE variation provides a means to alter the affinities of regulatory proteins to CRE sequences thereby modulating regulation (42, 43).

16.3.1.2 Bioenergetics

The requirement for high level cellulase synthesis in order for microbes to grow on cellulose raises the question: how does an anaerobic cellulolytic bacterium generate enough ATP for cellulase synthesis? We determined that *C. thermocellum* assimilated cellodextrins with a mean degree of polymerization (DP) of about four during growth on cellulose and that these cellodextrins were subsequently cleaved by intracellular phosphorolytic enzymes (44). The process is presented in Figure 16.4.

Soluble cellodextrins (G_n) from cellulose hydrolysis were transported across the cell membrane via the ABC transportation system with one ATP expenditure per molecule (45),



Although the sugar ABC transportation seems bioenergetically costly, it could be very important for (thermophilic) bacteria competing for low concentrations of hydrolysis products. These ABC transportation systems are widely observed in thermophilic or hyperthermophilic microorganisms with very high affinities of sugar uptake (46–48).

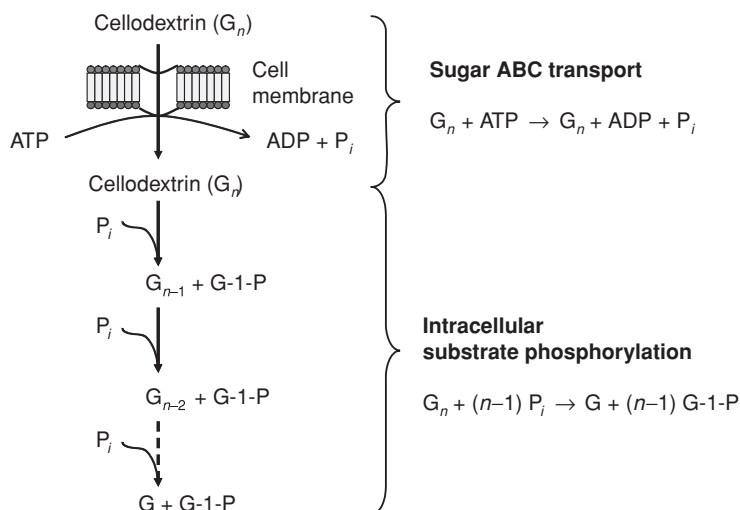
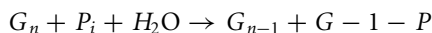


Figure 16.4 Cellodextrin transport and intracellular phosphorylation for *C. thermocellum*.

After sugar assimilation, intracellular cellodextrin and cellobiose phosphorylases, rather than β -glucosidase, cleave the β -(1-4) bonds of cellodextrins via substrate phosphorylation (49),



This phosphorylation process is important because it conserves energy stored in β -glucosidic bonds to generate one molecule of G-1-P (convertible to ATP) per cleavage and avoids energy dissipation resulting from hydrolysis mediated by β -glucosidase. Our bioenergetic analysis clearly indicated that by assimilating cellodextrins of average DP of 4, *C. thermocellum* has more ATP available for cell synthesis when growing on cellulose than on cellobiose (44). Hydrolysis products longer than cellobiose are supported by the supramolecular structure of the cellulosome and the distance between two adjacent catalytic subunits in the *C. thermocellum* cellulosome has been estimated to be eight glucosidic bonds (30). Thus, simultaneous catalytic events mediated by adjacent catalytic components would be expected to result in an insoluble G8 fragment, and any subsequent inter-cleavage of this G8 fragment would result in two soluble products with mean chain length 4. Low solubility (50) and/or a tendency to bind to cellulose may well prevent yet longer cellodextrins from being assimilated, even though this would in theory offer bioenergetic benefits. In summary, assimilation of longer cellodextrins means more energy generated from substrate phosphorylation and less energy expenditure for sugar transportation.

16.3.1.3 Enzyme–microbe synergy

Cellulose hydrolysis can be accomplished by cellulase enzymes acting without cells (e.g., cellulose hydrolysis in SHF), by cellulases acting in the presence of cells but with no cell–enzyme attachment (e.g., SSF), or by cellulases attached to cells (e.g., *C. thermocellum* in CBP). In the latter case, hydrolysis is mediated by ternary cellulose–enzyme–microbe (CEM) complexes rather than binary cellulose–enzyme (CE) complexes. For anaerobic cellulolytic bacteria, CEM complexes are commonly formed and are thought to be the major agent of cellulose hydrolysis (15). Potential benefits of CEM complexes for cellulolytic microorganisms have been suggested, including preferred access to hydrolysis products and local concentration of cellulases (15, 51–54).

Lu and coworkers (55) investigated such “cell–enzyme synergy” for *C. thermocellum*. Figure 16.5 shows that cellulose concentration changes are plotted against time for different cases:

- Microbial hydrolysis (CEM) – *C. thermocellum* culture with 100 mg/L cellulosome, 264 mg/L cell protein;
- Microbial control (CEM but non-active cells by chemical inhibitor) – a *C. thermocellum* culture with 100 mg/L cellulosome, 264 mg/L cell protein plus 38.5 mM sodium azide, a cell-growth inhibitor;
- CE in SSF – 100 mg/L purified *C. thermocellum* cellulase and active *T. saccharolyticum* that assimilates hydrolysis products;
- Cell-free control 1 (CE in SHF), 100 mg/L purified *C. thermocellum* cellulase; and

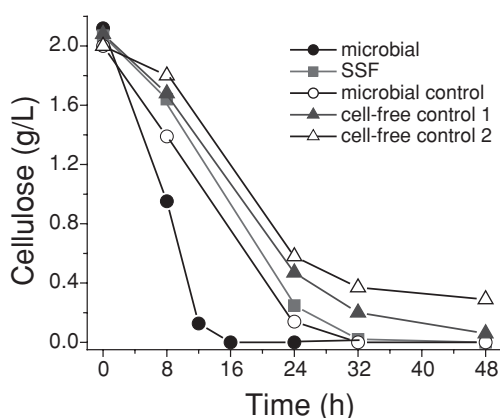


Figure 16.5 Comparison of *C. thermocellum* batch cellulose hydrolysis for microbial, SSF, and cellulase hydrolysis. [Data are redrawn from Lu *et al.* (55).]

- Cell-free control 2 (CE in SHF plus chemical inhibitor), 100 mg/L purified cellulase *C. thermocellum* cellulase with 38.5 mM sodium azide.

Comparison of microbial cultures (CEM) and cellulase–enzyme in the presence of *T. saccharolyticum* (SSF) cultures suggests that the *C. thermocellum* cellulase complex is substantially more effective during microbial hydrolysis compared to SSF under the conditions examined. Such “enzyme–microbe synergy” requires the presence of metabolically active cellulolytic microbes, and is not explained by removal of hydrolysis products from the bulk fermentation broth because (see control experiments).

From an applied perspective, the 2.7- to 4.7-fold synergistic effect reported is significant in the context of the search for strategies to decrease the cost of enzymatic hydrolysis, a focus of considerable effort since the late 1990s (3, 24, 56, 57). From a fundamental perspective, such synergy is one of the important inherent mechanisms to reduce cellulase synthesis requirement for anaerobic cultures.

16.3.1.4 Transformation systems

Gene transfer in Gram-positive bacteria can be based on natural competence (in rare instances), conjugation, transformation of protoplasts, or, most commonly, electrotransformation (ET) (58). Examples of gene transfer to Gram-positive obligate anaerobic thermophiles are limited to *Thermoanaerobacterium* species (59, 60) and a preliminary report of protoplast transformation of *C. thermocellum*, where the transformation efficiency was relatively low, e.g., $\sim 10^3/\mu\text{g}$ of DNA (61).

Tyurin and coworkers (58) conducted ET of several strains of *C. thermocellum* using plasmid pIKm1, developed by Mai and Wiegel (59, 60), as a transformation vector. A custom-built pulse generator was used to apply a square 10-ms pulse to an ET cuvette. Optimizations, including optimized periods of incubation both before and after electric pulse application, subculture in the presence of isoniazin prior to electric pulse application, use of a custom-built cuvette embedded in an ice block during pulse application, and high

(25-kV/cm) field strength, increased transformation efficiencies of *C. thermocellum* strain DSM 1313 from <1 to 2.2×10^5 transformants/ μg of DNA. The protocol and preferred conditions for strain DSM 1313 resulted in transformation efficiencies of 5.0×10^4 transformants/ μg of DNA for *C. thermocellum* strain ATCC 27405 and approximately 1×10^3 transformants/ μg of DNA for *C. thermocellum* strains DSM 4150 and 7072.

An interesting phenomenon was that current oscillations at about 24 MHz was observed during ET of Gram-positive bacteria *C. thermocellum* ATCC 27405, *C. thermocellum* DSM 1313, and *T. saccharolyticum* YS 485 using a pulse gated by a square signal generated by a custom generator (62). Oscillations were not observed for controls – including *E. coli* cells and a water blank. When a passive electrical filter consisting of an inductor and resistor in parallel was added to the system to prevent the development of oscillations, ET efficiencies were reduced dramatically for all three strains at all field strengths tested.

16.3.2 Recombinant cellulolytic strategy

The recombinant cellulolytic strategy involves engineering non-cellulolytic organisms that exhibit high product yields by producing a heterologous cellulase system enabling cellulose utilization. Early research advances have been reviewed previously (15). The yeast, *S. cerevisiae*, is a promising host organism for this strategy because of its high ethanol productivity at high yields, high osmo- and ethanol-tolerance, natural robustness in industrial processes, ease of genetic manipulation, and generally regarded as safe status due to its long association with the food and beverage industries. Cellulases from bacterial and fungal sources have been transferred to *S. cerevisiae*, enabling the hydrolysis of cellulosic derivatives (15), or growth on cellobiose (63, 64).

Three recombinant enzymes – *Trichoderma reesei* endoglucanase II, *T. reesei* cellobiohydrolase II, as well as *Aspergillus aculeatus* β -glucosidase cellulase – have been co-expressed in *S. cerevisiae* via individual fusion proteins with the C-terminal-half region of α -agglutinin (65). However, this recombinant strain cannot grow on cellulose using these recombinant cellulases, possibly because of poor recombinant cellulase expression or low enzyme activity or both.

van Zyl and coworkers (66) were the first to produce recombinant *S. cerevisiae* that can grow on pure insoluble cellulose by expressing two recombinant cellulases – the *T. reesei* endoglucanase (EG I) and the *S. fibuligera* β -glucosidase (BGL 1). The resulting strain was able to grow on phosphoric acid swollen cellulose (PASC) through simultaneous production of sufficient extracellular endoglucanase and β -glucosidase activity. Anaerobic growth was observed on the medium containing 10 g/L PASC as sole carbohydrate source with concomitant ethanol production of up to 1.0 g/L. Since crystalline cellulose hydrolysis requires three types of cellulases (endoglucanase, cellobiohydrolase, and β -glucosidase) to work together (24, 25), it is still a challenge to develop recombinant cellulolytic microorganisms that can express high levels of these cellulases to support cell growth on crystalline cellulose.

To achieve the self-supporting growth based on recombinant cellulases, it is appropriate to estimate the feasibility of cellulase expression levels. On the basis of the sufficiency of expression of growth-enabling heterologous enzymes, the level of enzyme expression required to achieve a specified growth rate may be calculated as a function of enzyme-specific activity (63). For growth enabled by cellulases with specific activities in the range available, required

expression levels are well within the range reported in the literature (1–10% of cellular protein) (67, 68). Protein expression at this level has been reported in both *S. cerevisiae* (67) and *E. coli* (69), although not to date for active cellulases. On the other hand, nature has created a diversity of cellulolytic microorganisms. With time, we anticipate that recombinant cellulolytic microorganisms with activity on crystalline cellulose will be created in the laboratory.

16.4 Future directions

The native cellulolytic organisms, e.g., *C. thermocellum*, are a starting point for developing anaerobic microbes capable of one-step processing of cellulosic biomass to ethanol or other desired products in the absence of added saccharolytic enzymes (12, 15, 28). However, it is necessary to perform metabolic engineering to improve product yield and titer relative to performance obtained to date with traditional ethanologenic strains. Substantial advances have been made in developing the genetic transformation tools for cellulolytic bacteria (58, 62, 70, 71). Recently, metabolic engineering has been applied to a mesophilic, cellulolytic *C. cellulolyticum* to achieve high ethanol yields. This was achieved by heterologous expression of *Zymomonas mobilis* pyruvate decarboxylase and alcohol dehydrogenase (70). The fermentation pattern was shifted significantly in that ethanol production increased by 53%, acetate increased by 93%, and lactate decreased by 48%.

From a fundamental viewpoint, it is also interesting to study the mechanisms of hydrolysis product uptake by natural cellulolytic microorganisms, determine hydrolysis product concentrations on the cell surfaces, and study the relationships between sugar transportation, energy level, and cellulase regulation.

From a practical viewpoint, another concern about naturally cellulolytic microorganisms is their sensitivity to the final product accumulation. For example, the wild *C. thermocellum* strain cannot grow in the presence of ethanol concentrations above 1% (w/v), yet industrial yeasts can tolerate ethanol titers up to 10% (w/v). Currently, an ethanol-adapted *C. thermocellum* strain that can tolerate up to 5.0–8.0% (w/v) ethanol was selected by step-wise increase in ethanol concentration (72). Another new method is global transcription machinery engineering (gTME), which re-programs gene transcription to tolerate substrate and product inhibition (73). Using gTME, a recombinant *S. cerevisiae* increased volumetric ethanol production by 69%, specific productivity by 41%, and ethanol yield by 14% (73). We expect that this technology could be applied to natural cellulolytic microorganisms as well.

The key objective for effective, recombinant cellulolytic ethanologens, e.g., *S. cerevisiae*, is the introduction of an efficient cellulase enzyme system. Key objectives to enable CBP microorganisms that we learned from natural cellulolytic microorganisms are to: 1) increase active recombinant cellulase expression levels up to 2–20% of cellular proteins (28), 2) optimize the expressed cellulase ratio (endo-/exo-/BG) by gene regulation, depending on their turnover number and substrate characterization (74), 3) generate more ATP from intracellular substrate phosphorylation by introduction of recombinant cellobiose and cel-lulodextrin phosphorylases, especially for anaerobic fermentations (44, 49), and 4) form an enzyme–microbe complex to reduce the enzyme synthesis burden by taking advantage of enzyme–microbe synergy (55).

Acknowledgment

Y.H.P. Zhang wishes to thank the Biological Systems Engineering Department of Virginia Polytechnic Institute and State University. Funding for this work also came from the USDA-CSREES 2006-38909-03484 grant and the DOE BSE and NIST.

References

1. Hoffert, M.I., Caldeira, K., Benford, G., Criswell, D.R., Green, C., Herzog, H., Jain, A.K., Kheshgi, H.S., Lackner, K.S., Lewis, J.S., Lightfoot, H.D., Manheimer, W., Mankins, J.C., Mauel, M.E., Perkins, L.J., Schlesinger, M.E., Volk, T.A. & Wigley, T.M. (2002) Advanced technology paths to global climate stability: Energy for a greenhouse planet. *Science*, **298**, 981–987.
2. Hoffert, M.I., Caldeira, K., Jain, A.K., Haites, E.F., Harvey, L.D.D., Potter, S.D., Schlesinger, M.E., Schneider, S.H., Watts, R.G., Wigley, T.M. & Wuebbles, D.J. (1998) Energy implications of future stabilization of atmospheric CO₂ content. *Nature*, **395**, 881–884.
3. Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J., Jr., Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R. & Tschaplinski, T. (2006) The path forward for biofuels and biomaterials. *Science*, **311**, 484–489.
4. Jarvis, M. (2003) Cellulose stacks up. *Science*, **426**, 611–612.
5. Berner, R.A. (2003) The long-term carbon cycle, fossil fuels and atmospheric composition. *Nature*, **426**, 323–326.
6. Cox, P.M., Betts, R.A., Jones, C.D., Spall, S.A. & Totterdell, I.J. (2000) Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, **408**, 184–187.
7. Falkowski, P., Scholes, R.J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F.T., Moore, B., III, Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V. & Steffen, W. (2000) The global carbon cycle: A test of our knowledge of earth as a system. *Science*, **290**, 291–296.
8. Melillo, J.M., Steudler, P.A., Aber, J.D., Newkirk, K., Lux, H., Bowles, F.P., Catricala, C., Magill, A., Ahrens, T. & Morrisseau, S. (2002) Soil warming and carbon-cycle feedbacks to the climate system. *Science*, **298**, 2173–2176.
9. Angenent, L.T., Karim, K., Al-Dahhan, M.H., Wrenn, B.A. & Domiguez-Espinosa, R. (2004) Production of bioenergy and biochemicals from industrial and agricultural wastewater. *Trends in Biotechnology*, **22**, 477–485.
10. Lynd, L.R., Wyman, C.E. & Gerngross, T.U. (1999) Biocommodity engineering. *Biotechnology Progress*, **15**, 777–793.
11. Wyman, C.E. (1999) Biomass ethanol: Technical progress, opportunities, and commercial challenges. *Annual Review Energy and Environment*, **24**, 189–226.
12. Demain, A.L., Newcomb, M. & Wu, J.H.D. (2005) Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews*, **69**, 124–154.
13. Lynd, L.R. (1996) Overview and evaluation of fuel ethanol from cellulosic biomass: Technology, economics, the environment, and policy. *Annual Review of Energy and Environment*, **21**, 403–465.
14. Lynd, L.R., Cushman, J.H., Nichols, R.J. & Wyman, C.E. (1991) Fuel ethanol from cellulosic biomass. *Science*, **251**, 1318–1323.
15. Lynd, L.R., Weimer, P.J., van Zyl, W.H. & Pretorius, I.S. (2002) Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, **66**, 506–577.
16. Kamm, B. & Kamm, M. (2004) Principles of biorefineries. *Application of Microbiology and Biotechnology*, **64**, 137–145.

17. Caldeira, K., Jain, A.K. & Hoffert, M.I. (2003) Climate sensitivity uncertainty and the need for energy without CO₂ emission. *Science*, **299**, 2052–2054.
18. Wirth, T.E., Gray, C.B. & Podesta, J.D. (2003) The future of energy policy. *Foreign Affairs*, **82**, 125–144.
19. Bush, G.W. (2006) *State of the Union Address by the President*; White House: January 31, 2006. Available from: <http://www.whitehouse.gov/news/releases/2006/01/20060131-10.html>.
20. Office of Biomass Program of DOE. (2004) *A Biomass Multi-Year Technical Plan: A Strong Energy Portfolio for a Strong America*. Available from: http://www.eere.energy.gov/biomass/pdfs/mytpsummary_040804.pdf.
21. Office of Energy Efficiency and Renewable Energy; Office of Science of DOE. (2006) *Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda*. A Research Roadmap Resulting from the Biomass to Biofuels Workshop. Available from: <http://www.doeenometolife.org/biofuels/>.
22. Wyman, C.E., Dale, B.E., Elander, R.T., Holtzapple, M., Ladisch, M.R. & Lee, Y.Y. (2005) Coordinated development of leading biomass pretreatment technologies. *Bioresource Technology*, **96**, 1959–1966.
23. Wyman, C.E., Dale, B.E., Elander, R.T., Holtzapple, M., Ladisch, M.R. & Lee, Y.Y. (2005) Comparative sugar recovery data from laboratory scale application of leading pretreatment technologies to corn stover. *Bioresource Technology*, **96**, 2026–2032.
24. Zhang, Y.-H.P., Himmel, M. & Mielenz, J.R. (2006) Outlook for cellulase improvement: Screening and selection strategies. *Biotechnology Advances*, **24**(5), 452–481.
25. Zhang, Y.-H.P. & Lynd, L.R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnology and Bioengineering*, **88**, 797–824.
26. Mosier, N., Wyman, C.E., Dale, B.E., Elander, R.T., Lee, Y.Y., Holtzapple, M. & Ladisch, M. (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technology*, **96**, 673–686.
27. Zhang, Y.-H.P., Ding, S.-Y., Mielenz, J.R., Elander, R., Laser, M., Himmel, M., McMillan, J.D. & Lynd, L.R. (2007) Fractionating recalcitrant lignocellulose at modest reaction conditions. *Biotechnology and Bioengineering*, **97**(2), 214–223.
28. Lynd, L.R., van Zyl, W.H., McBride, J.E. & Laser, M. (2005) Consolidated bioprocessing of cellulosic biomass: An update. *Current Opinion in Biotechnology*, **16**, 577–583.
29. DOE Biomass Program. (2004) *Cellulase Cost-Reduction Contracts*. Available from: http://www1.eere.energy.gov/biomass/cellulase_cost.html.
30. Mayer, F., Coughlan, M.P., Mori, Y. & Ljungdahl, L.G. (1987) Macromolecular organization of the cellulolytic enzyme complex of *Clostridium thermocellum* as revealed by electron microscopy. *Application of Environmental Microbiology*, **53**, 2785–2792.
31. Zhang, Y.-H.P. & Lynd, L.R. (2003) Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: Development of an ELISA-based method with application to *Clostridium thermocellum* batch cultures. *Analytical Chemistry*, **75**, 219–227.
32. Zhang, Y.-H.P. & Lynd, L.R. (2005) Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum*. *Journal of Bacteriology*, **187**, 99–106.
33. Bayer, E.A., Kenig, R. & Lamed, R.L. (1983) Adherence of *Clostridium thermocellum* to cellulose. *Journal of Bacteriology*, **156**, 818–827.
34. van Walsum, G.P. & Lynd, L.R. (1998) Allocation of ATP to synthesis of cells and hydrolytic enzymes in cellulolytic fermentative microorganisms: Bioenergetics, kinetics, and bioprocessing. *Biotechnology and Bioengineering*, **58**, 316–320.
35. Lynd, L.R. & Zhang, Y.H. (2002) Quantitative determination of cellulase concentration as distinct from cell concentration in studies of microbial cellulose utilization: Analytical framework and methodological approach. *Biotechnology and Bioengineering*, **77**, 467–475.
36. Johnson, E.A., Bouchot, F. & Demain, A.L. (1985) Regulation of cellulase formation in *Clostridium thermocellum*. *Journal of General Microbiology*, **131**, 2303–2308.

37. Mishra, S., Beguin, P. & Aubert, J.-P. (1991) Transcription of *Clostridium thermocellum* endoglucanases genes *celF* and *celD*. *Journal of Bacteriology*, **173**, 80–85.
38. Dror, T.W., Rolider, A., Bayer, E.A., Lamed, R.L. & Shoham, Y. (2003) Regulation of expression of scaffoldin-related genes in *Clostridium thermocellum*. *Journal of Bacteriology*, **185**, 5109–5116.
39. Dror, T.W., Morag, E., Rolider, A., Bayer, E.A., Lamed, R. & Shoham, Y. (2003) Regulation of the cellulosomal CelS (*cel48A*) gene of *Clostridium thermocellum* is growth rate dependent. *Journal of Bacteriology*, **185**, 3042–3048.
40. Stevenson, D.M. & Weimer, P.J. (2005) Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or Cellobiose in continuous culture. *Application of Environmental Microbiology*, **71**, 4672–4678.
41. Chaveaux, S. (1996) CcpA and HPr(ser-P): Mediators of catabolite repression in *Bacillus subtilis*. *Research in Microbiology*, **147**, 518–522.
42. Moreno, M.S., Schneider, B.L., Maile, R.R., Weyler, W. & Saier, J.M.H. (2001) Catabolite repression mediated by the CcpA protein in *Bacillus subtilis*: Novel modes of regulation revealed by whole-genome analysis. *Molecular Microbiology*, **39**, 1366–1381.
43. Gosseringer, R., Kuster, E., Galinier, A., Deutscher, J. & Hillen, W. (1997) Cooperative and non-cooperative DNA binding modes of catabolite control protein CcpA from *Bacillus megaterium* result from sensing two different signals. *Journal of Molecular Biology*, **266**, 665–676.
44. Zhang, Y.-H.P. & Lynd, L.R. (2005) Cellulose utilization by *Clostridium thermocellum*: Bioenergetics and hydrolysis product assimilation. *Proceedings of the National Academy of Sciences, USA*, **102**, 7321–7325.
45. Strobel, H.J., Caldwell, F.C. & Dawson, K.A. (1995) Carbohydrate transport by the anaerobic thermophilic *Clostridium thermocellum* LQRI. *Application of Environmental Microbiology*, **61**, 4012–4015.
46. Koning, S.M., Elferink, M.G.L., Konings, W.N. & Driessen, A.J.M. (2001) Cellobiose uptake in the hyperthermophilic Archaeon *Pyrococcus furiosus* is mediated by an inducible, high-affinity ABC transporter. *Journal of Bacteriology*, **183**, 4979–4984.
47. van Wely, K.H.M., Swaving, J., Freudl, R. & Driessen, A.J.M. (2001) Translocation of proteins across the cell envelope of Gram-positive bacteria. *FEMS Microbiology Reviews*, **25**, 437–454.
48. Elferink, M.G.L., Albers, S.-V., Konings, W.N. & Driessen, A.J.M. (2001) Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. *Molecular Microbiology*, **39**, 1494–1503.
49. Zhang, Y.-H.P. & Lynd, L.R. (2004) Kinetics and relative importance of phosphorylytic and hydrolytic cleavage of cellodextrins and cellobiose in cell extracts of *Clostridium thermocellum*. *Application of Environmental Microbiology*, **70**, 1563–1569.
50. Zhang, Y.-H.P. & Lynd, L.R. (2003) Cellodextrin preparation by mixed-acid hydrolysis and chromatographic separation. *Analytical Biochemistry*, **322**, 225–232.
51. Adams, J.J., Pal, G., Jia, Z. & Smith, S.P. (2006) Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proceedings of the National Academy of Sciences, U.S.A.*, **103**(2), 305–310.
52. Miron, J., Ben-Ghedalia, D. & Morrison, M. (2001) Invited review: Adhesion mechanisms of rumen cellulolytic bacteria. *Journal of Dairy Science*, **84**, 1294–1309.
53. Schwarz, W.H. (2001) The cellulosome and cellulose degradation by anaerobic bacteria. *Application of Microbiology and Biotechnology*, **56**(5–6), 634–649.
54. Shoham, Y., Lamed, R. & Bayer, E.A. (1999) The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends in Microbiology*, **7**(7), 275–281.
55. Lu, Y., Zhang, Y.-H.P. & Lynd, L.R. (2006) Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *Proceedings of the National Academy of Sciences, USA*, **103**(44), 16165–16169.

56. Knauf, M. & Moniruzzaman, M. (2004) Lignocellulosic biomass processing: A perspective. *International Sugar Journal*, **106**, 147–150.
57. Sheehan, J. & Himmel, M. (1999) Enzymes, energy, and the environment: A strategic perspective on the U.S. department of energy's research and development activities for bioethanol. *Biotechnology Progress*, **15**, 817–827.
58. Tyurin, M.V., Desai, S.G. & Lynd, L.R. (2004) Electrotransformation of *Clostridium thermocellum*. *Application of Environmental Microbiology*, **70**, 883–890.
59. Mai, V., Lorenz, W.W. & Wiegel, J. (1997) Transformation of *Thermoanaerobacterium* sp. strain JW/SL-YS485 with plasmid pIKM1 conferring kanamycin resistance. *FEMS Microbiology Letters*, **148**, 163–167.
60. Mai, V. & Wiegel, J. (2000) Advances in development of a genetics system for *Thermoanaerobacterium* spp. expression of genes encoding hydrolytic enzymes, development of a second shuttle vector, and integration of genes into the chromosome. *Application of Environmental Microbiology*, **66**, 4817–4821.
61. Tsoi, T.V., Chuvil'skaya, N.A., Atakishieva, Y.Y., Dzhavakhishvili, T.D., Akimenko, V.K. & Boronin, A.M. (1987) *Clostridium thermocellum* – a new object of genetic investigations. *Molekuliarnaya Genetika, Mikrobiologiya, I Virusologiya*, **11**, 18–23.
62. Tyurin, M.V., Sullivan, C.R. & Lynd, L.R. (2005) Role of spontaneous current oscillations during high-efficiency electrotransformation of thermophilic anaerobes. *Application of Environmental Microbiology*, **71**(12), 8069–8076.
63. McBride, J.E., Zietsman, J.J., Van Zyl, W.H. & Lynd, L.R. (2005) Utilization of cellobiose by recombinant beta-glucosidase-expressing strains of *Saccharomyces cerevisiae*: Characterization and evaluation of the sufficiency of expression. *Enzyme and Microbial Technology*, **37**(1), 93–101.
64. van Rooyen, R., Hahn-Hagerdal, B., La Grange, D.C. & van Zyl, W.H. (2005) Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. *Journal of Biotechnology*, **120**(3), 284–295.
65. Fujita, Y., Ito, J., Ueda, M., Fukuda, H. & Kondo, A. (2004) Synergistic saccharification, and direct fermentation to ethanol, of amorphous cellulose by use of an engineered yeast strain codisplaying three types of cellulolytic enzyme. *Application of Environmental Microbiology*, **70**, 1207–1212.
66. Den Haan, R., Rose, S.H., Lynd, L.R. & van Zyl, W.H. (2007) Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. *Metabolic Engineering*, **9**, 87–94.
67. Park, E.-H., Shin, Y.-M., Lim, Y.-Y., Kwon, T.-H., Kim, D.-H. & Yang, M.-S. (2000) Expression of glucose oxidase by using recombinant yeast. *Journal of Biotechnology*, **81**(1), 35–44.
68. Romanos, M.A., Makoff, A.J., Fairweather, N.F., Beesley, K.M., Slater, D.E., Rayment, F.B., Payne, M.M. & Clare, J.J. (1991) Expression of tetanus toxin fragment C in yeast: Gene synthesis is required to eliminate fortuitous polyadenylation sites in AT-rich DNA. *Nucleic Acids Research*, **19**(7), 1461–1467.
69. Hasenwinkle, D., Jervis, E., Kops, O., Liu, C., Lesnicki, G., Haynes, C.A. & Kilburn, D.G. (1997) Very high-level production and export in *Escherichia coli* of a cellulose binding domain for use in a generic secretion-affinity fusion system. *Biotechnology and Bioengineering*, **55**(6), 854–863.
70. Guedon, E., Desvaux, M. & Petitdemange, H. (2002) Improvement of cellulolytic properties of *Clostridium cellulolyticum* by metabolic engineering. *Application of Environmental Microbiology*, **68**, 53–58.
71. Jennert, K.C.B., Tardif, C., Young, D.I. & Young, M. (2000) Gene transfer to *Clostridium cellulolyticum* ATCC 35319. *Microbiology*, **146**, 3071–3080.

72. Taufika Islam, W., Combs, J.C., Lynn, B.C. & Strobel, H.J. (2007) Proteomic profile changes in membranes of ethanol-tolerant *Clostridium thermocellum*. *Application of Microbiology and Biotechnology*, **74**, 422–432.
73. Alper, H., Moxley, J., Nevoigt, E., Fink, G.R. & Stephanopoulos, G. (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science*, **314**(5805), 1565–1568.
74. Zhang, Y.-H.P. & Lynd, L.R. (2006) A functionally-based model for hydrolysis of cellulose by fungal cellulase. *Biotechnology and Bioengineering*, **94**(5), 888–898.

Index

- 4-CL modulation, 252–4
- 4-coumarate CoA ligases, 229–30
- α -arabinofuranosidase, 363
- accessory enzymes, 362–4
- Acetobacter xylinum*, 190
- acetyl bromide, 239–40
- acetyl xylan esterase, 363–4
- acetyltransferases, 119–220
- acid hydrolysis, 17, 331
- acid mine drainage (AMD) biofilm, 468
- Acidothermus cellulolyticus*, 386
 - colonized, 459f
- acinobacteria, 461
- acylated homoserine lactones (AHLs), 461
- Adler model, 272–3
- ADP glucose, 145–6
- aerobic fungi, 458
 - cellulolytic, 375
- aerobic microbial cellulase systems, 374–5
- AFEX. *See* ammonia, fiber explosion
- AFM. *See* atomic force microscope
- aggregation
 - alternative patterns, 203–10
 - isolation changes, 206–10
 - levels, 209f
 - patterns, 209f
- AHLs. *See* acylated homoserine lactones
- alcohol conformation nomenclature, 195f
- alfalfa pC3H line, 250
- aliphatic polyester, 78
- alkaline nitrobenzene oxidation (NBO), 240–41
- alkaline pulping delignification, 237
- AMBER force field, 308–10, 313
- AMD. *See* acid mine drainage
- American Type Culture Collection, 462
- ammonia
 - fiber explosion (AFEX), 54, 353
 - pretreatment, 441–2, 444–6
- ammonia-recycled percolation (ARP), 444–5
- anaerobic bacteria, clostridium-related, 396
- anaerobic rumen fungus, 377
- analysis methods, 317–19
- angiosperms, 220
 - herbaceous, 62
- anhydroglucose repeat periods, structure loci, 198f
- AP:ApiT. *See*
 - apiogalacturonan-apiosyltransferase
- AP:GalAT. *See* apiogalacturonan-galacturonosyltransferase
- apiogalacturonan synthesis, 127–8
- apiogalacturonan-apiosyltransferase (AP:ApiT), 127–8
- apiogalacturonan-galacturonosyltransferase (AP:GalAT), 127–8
- Arabidopsis*, 228–30
 - CAD mutation and vascular integrity, 259–60
 - COMT-downregulated line, 263
 - tissue analysis, 102
- Arabidopsis thaliana*
 - cell wall, 42
 - vascular gene expression, 223f
- arabino-3,6-galactans, 69–70
- Arabinogalactan, 366–7
- arabinoxylan chains, cross-linked, 73f
- ARISA. *See* rRNA, automated
- Army Quartermaster Lab (Natick, MA), 381
- arogenate, biosynthetic pathway, 226f
- ARP. *See* ammonia-recycled percolation
- Ascomycota, xylariaceous, 456
- aspen, 253–4
- Atalla laboratory, 207

- atomic force microscope (AFM), 49–50
 imaging, 188, 192*f*
Atomic Force Microscopy, 319
automated RISA (ARISA), 467
autohydrolysis, 442
- β -glucan, 366–7
 β -glucanses, 361
 cell wall depolymerizing, 356*t*
 cell wall polysaccharide-depolymerizing, 356*t*
Bacillus agaradhaerens, 386
bacteria
 cellulolytic aerobic, 382–6
 saprophytic fungi and, 463–4
bacterial cellulose ribbons, degradation, 419*f*, 421*f*
bacterial genome, dockerin-like sequence, 415
bacterial microcrystalline cellulose (BMCC), 419
bacterial producers, 461
Bacteroides species, 397
barrier reaction
 energies, 343*f*
 absence, 343
Basidiomycota, white rot, 456
Beechwood xylan slurries, decomposition, 346*f*
bend energy, 308–9
benzyl ether linkages, 67*f*, 76–7
biochemical conversion, 10*f*, 17–23, 18*f*, 279*f*
biodiversity estimates, 455
bioenergetics, 485–6
bioenergy plant engineering, objective, 2
bioengineering research, 32
biofuel production potential, 7–8, 8*f*
biological pretreatment, 447
biological science, R&D role, 30–32
biomass
 current conversion technology, 53
 enzymes, 437
 fractionation, 34
 gasification process, 24
 microarray methods sampling, 470–72
 microbial communities and, 464–70
 processing configurations, 482*f*
 production capability, 7
 recalcitrance, 1–2, 407–8
biomass decay
 communities, 456–7
 systems biology approach, 454–6
biomass-degrading communities
 defining, 465–7
 molecular approaches, 466–7
biomass-degrading enzymes, 2–3
biomass feedstock
 hemicellulase activities, 364–7
 preprocessing and blending technologies, 29
 biomass feedstock pretreatment, 439–49
 cellulose digestibility and, 438–9
 processes, 437–8
bio-oil, 11
biophotonics, 50
bioprocessing, consolidated, 481–3
biorefinery
 advanced, 28
 gasification, integrated scenario, 34*f*
 technology advancement, 28
biorefinery concept, stages, 9
biosynthesis, 42–5
BMCC. *See* bacterial microcrystalline cellulose
Boltzman's constant, 322
Born-Oppenheimer approximation, 307
BRENDA. *See* Comprehensive Enzyme Information System
- CADs. *See* cinnamyl alcohol dehydrogenases
caffeic acid *O*-methyltransferases (COMTs), 230–32
 activity, 262–5
 downregulation, 263
caffeoyl CoA *O*-methyltransferases (CCOMTs), 230–32
Carbohydrate Active enZymes (CAZy) database, 159, 354, 456
carbohydrate-binding module (CBM), 375, 408
carbohydrate force fields, 314
carbon
 catabolite repression (CCR), 484
 cation product, 341
 turnover, 455–6
carboxylic ester, 76*f*
carboxymethylcellulose (CMC), 375
Car-Parrinello molecular dynamics (CPMD), 333
CARS. *See* coherent anti-stokes Raman spectroscopy
catalytic constants, pH optimum comparison, 122*t*
catalytic domain (CD), 375
CAZy. *See* Carbohydrate Active enZymes database
CBHA. *See* *Clostridium thermocellum*
CBM. *See* carbohydrate-binding module
CBP. *See* consolidated bioprocessing
CBS. *See* complete basis set
CCOMTs. *See* caffeoyl CoA *O*-methyltransferases
CCR. *See* carbon, catabolite repression
CCRs. *See* cinnamoyl CoA reductases
CD. *See* catalytic domain
CE. *See* cellulose-enzyme complexes
cell biology, cell wall synthesis and, 136–7

- cellular location, 158
- cellulase
 - activities, 460
 - diversity, 376–9
 - vs. free cellulase systems, 423*f*
 - fungal, 381–2
 - genes, site-directed mutagenesis, 384
 - role, 375–6
 - regulation, 484–5
 - synergism, 380
 - synthesis, 383
 - types, 376–7
- cellulolytic anaerobic soil bacteria, 374
- cellulolytic bacteria, 400
- cellulolytic microorganism sensitivity, 489
- cellulose, 3–4, 48, 64–70, 323
 - chains, paracrystalline arrangement, 407
 - crystal structures, 48
 - crystal unit cell, 195*f*
 - deconstruction, 424
 - degradation, 320–23
 - deposition, 100–101
 - digestibility and pretreated biomass, 438–9
 - hydrolysis, 322–4, 486–7, 487*f*
 - microfibrils, 96–8, 190–94
 - modeling, 324–5
 - model, *T. fusca* Cel9A, 378*f*
 - native in living plants, 203
 - purification methods, 206
 - structure, 65*f*, 97, 208
- cellulose-binding domains, 379–80, 413
- cellulose-binding module (CBM), 408, 424–6
- cellulose-enzyme complexes (CE), 486
- cellulose-enzyme-microbe (CEM), 486
- cellulose synthase, 97
 - activity, 98–100
 - complex (CelS), 42
 - genes, 102–3
 - proteins (CesA), 42, 97, 99, 100–103
- cellulose synthase-like 4 (CSLC4), proteins, 108
- cellulose synthesis, 96–8
 - inhibition, 99
 - regulation, 101–4
 - R. fruticosus*, 98
- cellulosic biomass, conversion routes, 10*f*
- cellulosic feedstocks, organism development, 483*f*
- cellulosomal carbohydrate-active enzymes, 410–15
- cellulosome, 407–8
 - assembly, 412
 - bacteria interactions, 418*f*
 - binary designers, synergy of, 425*f*
 - cell-surface disposition, 417–18
 - cellulose binding, 416
 - cellulose interaction, 415–17
 - complex, 424
 - concept, 408–10
 - electron micrograph, 416*f*
 - organization, cell surface, 411*f*
 - rationale, 423–6
- cell walls
 - biosynthesis, 41–9
 - breakdown, 398–9
 - cellulose elementary fibril model, 43*f*
 - compositions, cellulosic, 63*t*
 - degradabilities by enzymes, 83–5
 - hemicelluloses in, 52
 - lamellae, 45–6
 - layers, 45*f*
 - modeling potentials, 325–6
 - molecular architecture, 79–83
 - molecular structure, 41–9
 - multi-scale, 325*f*
 - polymer chemistry, 62–70
 - polysaccharide chemistry, 62–70
 - pore system, 40–41
 - primary, 40, 79–81, 80*f*, 95*t*
 - protein chemistry, 70
 - saccharification, 31
 - secondary, 40, 95*t*
 - structure characterization, 49–53
 - sugars, 4–5
 - synthesis, cell biology and, 136–7
 - synthesis schematic, 44*f*
 - transcriptional control, 286
 - types, 61–2
- CelS. *See* cellulose synthase, complex
- CEM. *See* cellulose-enzyme-microbe
- CesA. *See* cellulose synthase, proteins
- CESA proteins, 97, 99, 100–103
- C4H downregulation/mutation, 244–8
- chain interactions, monolignol radical and lignin primary, 286
- chromatin immunoprecipitated DNA, 470
- chromophoric substrates, 266*f*
- cinnamoyl CoA reductases (CCRs), 230, 255–6
- cinnamyl alcohol dehydrogenases (CADs), 230, 257–60
 - mis-annotation, 260
- Clostridia, 413–14
- Clostridium acetobutylicum*, 425–6
- Clostridium cellulolyticum*, 461
- Clostridium thermocellum* (CBHA), 377, 408–10, 425–6, 461, 483, 485
 - bacterial cell surface attachment, 409*f*
- clustering, 318
- CMC. *See* carboxymethylcellulose
- CMP-KDO, 154–5

- coherent anti-stokes Raman spectroscopy (CARS), 50
cohesin-dockerin interaction, 408
comminution, 440
complete basis set (CBS), 334
Comprehensive Enzyme Information System (BRENDA), 354
compression/tension wood tissues, 223–5
computer simulations, 51–3
COMTs. *See* caffeic acid *O*-methyltransferases
consolidated bioprocessing (CBP), 482–9
corn stem, inoculated and colonized, 458*f*
corn stover, 347–8
 pretreated rind, 449*f*
cortical microtubules, 100
Coulomb's law, 310
covalent cross-linking, 78–9
 polysaccharide-lignin types, 74–7
CPMD. *See* Car-Parrinello molecular dynamics
cross-linking, 72–9
 covalent modes, 72*f*, 78–9
crystallinity, approaches, 210
CSLC4. *See* cellulose synthase-like 4
cutin, 78
cutoffs, non-bonded, 311–12
crystallographic study flaw, 189–90
Cytophaga hutchinsonii, 383–4, 463

debranching enzymes, 362–4
degree of polymerization (DP), 96, 485
dehydrogenative polymerization, 274–5
denaturing gradient gel electrophoresis (DGGE), 467, 468
Department of Energy (DOE) Joint Genome Institute, 382, 461
depolymerases, 359–62
DGGE. *See* denaturing gradient gel electrophoresis
DHPs. *See* synthetic dehydropolymerizates
diGalA. *See* digalacturonic acid
digalacturonic acid (diGalA), 121
dilute acid
 batch/co-current pretreatment, 443
 hydrolysis, xylem SEM imaging, 448*f*
 percolation/countercurrent pretreatment, 443–4
dimeric derivatives, 251*f*
dirigent proteins
 abundance, 280
 coupling, 278*f*
DNA-based fingerprinting techniques, 466
DOE. *See* Department of Energy
downregulation/mutation
 C4H, 244–8
 monolignol radical regeneration, 254–67
 PAL, C4H, and 4CL, 243–4
 pC3H and HCT, 243–4, 249–52
DP. *See* degree of polymerization
dynamics methods, 316–17

electron microscopy (EM), 49
electrostatics, long range, 311–12
EM. *See* electron microscopy
endocellulase, 377*f*
endoglucanases, 381
enhanced sampling methods, 319–22
enzymatic hydrolysis, 482*f*
enzyme-microbe synergy, 486–7, 489
enzymes
 accessory, 362–4
 plant cell-wall degrading, 414*t*
 production, R&D needs, 21–2
 topology, 158
enzymology, 98–100
ester
 -ester cross-links, 74
 linkages, direct, 74–5
ethanol
 cellulosic feedstocks, 63*t*
 gasoline price comparison, 24*f*
 prices, 23*f*
 production potential, 8*t*
eukaryotic microorganisms, 398
exocellulase
 classes, 376–7, 377*f*
 studies, 385–6
ExPASy. *See* Expert Protein Analysis System
Expert Protein Analysis System (ExPASy), 354
expression profiling, 469

feedstocks, 12–17
 costs, 480
 digestibility pretreatments, 436–7
 economic viability, 13*f*, 14–15
 engineering, 30
 preprocessing, 16, 29*f*
 prices, liquid transportation fuels, 481*f*
 queuing, 16–17
 R&D pathway, 13–14
 storage, 16–17
 supply system, 15–16, 28–9
 transportation and handling, 17
 wet and dry herbaceous, 14
fermentation
 history, 140–58
 R&D, toxicity and, 20
ferulate 5-hydroxylase (F5H), 260–62
ferulic acid esterase, 364
F5H. *See* ferulate 5-hydroxylase

- fibril, macro and elementary, 46–7
Fibrobacter, 400
 succinogenes, 384, 396–7
finite difference methods, 316
fluorescence protein (FP), 51
force field, 308, 313–15
 equation, 307–8
Forss lignin models, 269–72
Fourier-transform infrared (FTIR), 135
FP. *See* fluorescence protein
FRA8. *See* *FRAGILE FIBER8*
FRAGILE FIBER8 (*FRA8*), 108
free energy methods, 319–22
Freudenberg lignin models, 269–72
FTIR. *See* Fourier-transform infrared
fungal cellulases, 381–2
fungi, plant biomass and, 457–60
furfural formation, 347*f*
- α -D-galactosidase, 363
 α -glucuronidase, 362–3
 β -glucan, 366–7
 β -glucanses, 361
 cell wall depolymerizing, 356*t*
 cell wall polysaccharide-depolymerizing,
 355*t*
galactoglucomannans, 67, 366
galactosyltransferase (GalT) activities,
 132–3
GalATs, 121–3
 genes, 123
GalT. *See* galactosyltransferase
gasification, 33
 process, 25
gasoline, and ethanol price comparison, 24*f*
Gaussian programs, 334
GAUT genes, 123
GAUT1, 120–21
 arabidopsis, gene superfamily, 124*t*
 pectin synthesis function, 125
GAXs. *See* glucuronoarabinoxylans
GDP-Fuc, 153–4
GDP-Gal, 154
GDP-Man, 153
gene transfer, horizontal, 398
genome sequencing, organisms with
 cellulosomal components, 412
genomic signature tags (GST), paired-end,
 470
genomics revolution, 5
GFP. *See* green-fluorescence-protein
GH. *See* glycoside hydrolases
Ginkgo biloba, 2
Glasser and Glasser model, 272–3
Glaucocystis algae, 48
Golgi, 158
glucomannans, 67, 366
glucuronoarabinoxylans (GAXs), 110
 rich walls, 81
glucuronoxylan hydrolysis, deacetylated, 359*t*
glycoside hydrolases (GH), 54
glycosyl hydrolases, cell wall depolymerizing,
 356*t*
glycosyltransferases (GTs), 115
 gene candidates, 159
 mutations, 108
green-fluorescence-protein (GFP), 51
GST. *See* genomic signature tag
GTs. *See* glycosyltransferases
gymnosperm western red cedar, 284*f*
- Halocynthia*, 190
 roretzi, 48
HCA. *See* hydroxycinnamic acid
HCT silencing, 252
heartwood tissues, 223–5, 224*f*
Helmholtz free energy, 322
hemicelluloses, 52, 104
 activities, specificities, and types, 354–9
 enzymatic depolymerization, 352–3
 hydrolysis of solubilized, 367–8
herbaceous angiosperms, 62
heteroglucans, 64–6
heteromannans, 67
heteroxylans, 66–7
HG-acetyltransferase (HG-AT), 126
HG-AT. *See* HG-acetyltransferase
HG-galacturonosyltransferase (HG:GalAT),
 120–25
HG:GalAT. *See* HG-galacturonosyltransferase
HG-methyltransferase (HG-MT), 125–6
HG-MT. *See* HG-methyltransferase
high pressure liquid chromatography (HPLC),
 344
homogalacturonan synthesis, 120–26
hot spring samples, 461
HPLC. *See* high pressure liquid chromatography
HSF. *See* hybrid saccharification and
 co-fermentation
hybrid QM/MM potential, 317
hybrid saccharification (HSF) and
 co-fermentation, 18
hydrogen bonds
 inter-plane, 197*f*
 patterns, 196*f*
hydrolysis
 degradation, mechanisms, and kinetics, 333*f*
 investigation, 344
 structure and, 323–4
hydroxycinnamic acid (HCA), esterified, 71

- interactions
 - electrostatic, 310
 - non-bonded, 310
 - van der Waals, 311
- interatomic potentials, 308–11
- intermolecular coupling, 281*f*
- internal transcribed spacers (ITS), 467
- intro-molecular coupling, 281*f*
- irradiation, 440
- IRREGULAR XYLEM8 (IRX8)*, 108
- IRX8*. *See* *IRREGULAR XYLEM8*
- IRX9* gene mutation, 109
- isoform role, 156–7
- ITS. *See* internal transcribed spacers
- Klason methods, 239–40
- land resource potential, 8*f*
- large intestine, 400
- LCC. *See* lignin-carbohydrate complexes
- Lennard-Jones
 - function, 311
 - potential, 312
- lignans
 - racemic, 270*f*
 - structural depictions, 272–4
- lignification, 218–23, 268
 - cell wall formation, 225
 - protein role, 265–7
- lignified secondary walls, 81–3
- lignin analyses
 - shortcomings, 235–41
 - subunit and structural, 237–9, 238*f*
- lignin-carbohydrate complexes (LCC), 72
- lignin-degrading enzyme identification, 215
- lignin macromolecular configuration, 262–5, 268–9, 274–84
 - assembly, 285–7
- lignin pathway, vascular anatomical
 - development, 218–25
- lignin properties, 275–6
- lignins
 - abundance, 213, 215
 - biosynthesis, rate-limiting factor, 249
 - contents plotting, 256
 - crosslinks, 75*f*
 - degradation, 456
 - deposition, 221–2
 - initiation sites, 285
 - land adaptation, 213, 215
 - localization, 220–21
 - macromolecular configuration, 215
 - models, 269–72
 - molecular basis, 213, 215
 - physiological functions, 220
 - plant stem roles, 222
 - primitive plant chemistry, 220
 - quantification and degradation, 239–41
 - red, 257–8
 - synthetic dehydropolymerizates, 239–41
- lignin structures, 217*f*
 - depictions, 273–4
 - racemisation, 282–4
- lignin template
 - polymerization, 242–3
 - replication, 285–7
- lignin utilization, 33
- lignocellulose biorefineries, 1, 10–12, 54
 - biomass, pretreatment effect, 439*t*
 - breakdown, 400
 - degradation, ecology, 456
- lime pretreatment, 445
- linkages, interunit and benzodioxane, 264*f*
- liquid hot water
 - batch pretreatment, 442
 - percolation pretreatment, 442
- loblolly pine
 - CAD mutant, 258
 - tissue lignification and cell wall development, 221*f*
- logging residues, 15
- maize, 254
 - cell wall anatomy, 38–41
 - parenchyma cell wall, 52*f*
 - parenchyma surface structure, 46*f*
 - stem, transverse section, 39*f*
- mannan, 104–5
- mannanases, 360–61
- marine micro organisms, 463
- matrix polymers, 49
- MD. *See* molecular dynamics
- metabolic flux analyses, 225–30
- metabolism, monolignol to terpenoid, 276–84
- metagenomics, 399–400
 - analysis, 456, 468–9
- methoxyl group composition, invariant of*, 250
- methyltransferases (MetTs), 119, 137
- MetTs. *See* methyltransferases
- Micrasterias denticulata*, 190, 191*f*, 208
- microarray hybridization, DNA reference, 471*f*
- microbial communities, biomass and, 464–70
- microcrystalline cellulose, solubilization, 424*f*
- microfibril, 47–8
 - surface structure, 47*f*
- microtubule-cellulose synthase, 101
- microwave
 - energy, 344
 - reactor, constants and temperature, 338*f*
- mixed-linkage glucans (MLGs), 110

- MLGs. *See* mixed-linkage glucans
MM. *See* molecular mechanics
molecular architecture, 79–83
molecular dynamics (MD), 306, 315, 317
 simulations, 194, 333–4
 steered, 319
 targeted (TMD), 319
molecular mechanics (MM), 306–13
 potential, 317
molecular model types, 312–13
molecular modeling, 194–200
monolignol biosynthesis, 225–30
monolignol/lignin forming pathway
 (*Arabidopsis*)
 database, 234–5
 metabolic networks, 234–5
monolignol metabolism
 CAD1, 232–3
 proteins in, 232–4
 sinapyl alcohol dehydrogenase, 233–4
monolignol pathway
 genetic manipulation, 225
 modulation, 242–3
 metabolic flux analyses, 227
 transcriptional profiling, 227
monolignol radical-binding proteins, 276
monolignol radical generation, 265
 enigma of, 254–67
monolignol transport, 285
monomeric derivatives, 251*f*
morphology, normal vessel, 108
mung bean
 cell walls, 79
 microsomes, 133
myo-inositol pathway, 147–8

NADP⁺, substrate-binding pocket structure, 231*f*
nanofibrils
 aggregation, 205*f*
 geometric representation, 193*f*
National Park Service, 462
National Renewable Energy Laboratory (NREL), 9
native cellulolytic microorganisms, 483–8
native cellulose composites, 189
native lignin
 macromolecular configuration, 282–4
 protein directed, 282–4
NBO. *See* alkaline nitrobenzene oxidation
NDP. *See* nucleoside-diphosphate
NDP-sugars. *See* nucleotide sugars
NEB. *See* nudged elastic band
Nimz model, 272–3
NMR. *See* Nuclear Magnetic Resonance
nonlinear microscopy, 50
normal mode analysis, 318
North American Consortium for Genomics of Fibrolytic Ruminant Bacteria, 384
NREL. *See* National Renewable Energy Laboratory
Nuclear Magnetic Resonance (NMR)
 experiments, 319
nucleoside-diphosphate (NDP), 115
nucleotide sugars (NDP-sugars), 137–40
 biosynthetic genes, 138–9
 enzymes, 155–9
 history, 140–58
 interconversion pathway, 140–42
 metabolism, 141*f*
 pyrophosphorylase, 142–3
nudged elastic band (NEB), 320

one-step processing, 489
Ordovician period, 213
organic solvents pretreatment, 445–6
oxidation
 degradation, 240–41
 polymerization, proteins and, 267
 pretreatment, 446–7
oxonium ion, 332*f*

PA. *See* proton affinities
packing effects, 204
PAL. *See* phenylalanine ammonia lyase
parenchyma cell wall, 41*f*
Particle Mesh Ewald Method (PME), 312
PB. *See* Poisson-Boltzmann
pectic polysaccharides, 67–9
 structures, 111*f*, 113
pectin, 67–9, 110–14
 biosynthesis, glycosyltransferases required, 115, 116–18*t*
 modifying enzymes, 119–200
 schematic structure, 112*f*
pectinase activities, 460
pectin-rich walls, 79–80, 80*f*
pectin synthesis
 location, 114–15
 non-glycosyltransferases for, 119*t*
permanganate oxidation degradation, 240–41
peroxidase, 267
 modulation, 242–3
pH optimum, catalytic constants comparison, 121*t*
phase polysaccharides, microfibrillar phase, 64–70
phenolic coupling, 281*f*
 dirigent proteins in, 276–84
phenyl glycoside linkages, 77, 77*f*

- phenylalanine ammonia lyase (PAL), 227–8
 downregulation/mutation, 243–4
 inhibitor and phenolic-derived substructures, 245*f*, 246*f*, 247*n*
 inhibitor and substructures, 246*f*, 247*n*
- phenylalanine formation, 226–7
- phenylpropanoid genes
 changes and effects, 248*f*
 downregulating varieties, 248*f*
 pathway view, 214*f*
- phenylpropanoids, 216*f*
- Pichia pastoris*, 100
- pinoresinol-lariciresinol reductase protein, 284, 284*f*
- plant biomass
 bacteria identified, 460–463
 fungi and, 457–60
 substrate diversity, 455
- plant cell walls. *See* cell walls
- plant cellulose synthase, activated, 102
- plant lineages, structures in extant, 219*f*
- plant stem tissues, categorized, 39–40
- PME. *See* Particle Mesh Ewald Method
- PMF. *See* potential of mean force
- Poisson-Boltzmann (PB) solvers, 315
- polyaliphatic-polyphenolic association, 78
- polymerization
 degree of (DP), 96, 485
 dehydrogenative, 274–5
- polysaccharides
 associations, 71
 crosslinks, 75*f*
 debranching enzymes, cell wall, 357*t*
 degrading enzymes, 413
 lignin cross-linking, 72–4
 matrix phase, 64–70
 utilization, 78–9
- Populus alba* x *tremula*, 109
- potential of mean force (PMF), 322
- prephenate, biosynthetic pathway, 226*f*
- pretreatments
 acidic, 443–4
 AFEX, 441–2, 446
 alkaline, 444–5
 autohydrolysis, 442
 biological, 447
 biomass feedstock, 439–49
 oxidative, 446–7
 physical, 440
 prospects, 447–9
 rapid decompression, 440–42
 solvent, 445–6
 supercritical fluid, 446
- Prevotella* species, 397
- primary cell wall, pectin-rich polymer arrangement, 80*f*
- process optimization, 34–5
- protein roles
 lignification, 265–7
 oxidation/polymerization, 267
- proteins
 lignin cross-linking, 78
 monolignol and dirigent, 276–84
 polysaccharide cross-linking, 77–8
- proton affinities (PA), 338*f*
- putative arabinan, 134
- Pyrococcus horikoshii*, 386
- pyrolysis, 11, 33
- QM. *See* quantum mechanics
- QM/MM. *See* quantum mechanical/molecular mechanical
- QTLs. *See* quantitative trait loci
- quantitative trait loci (QTLs), 135
- quantum mechanical/molecular mechanical (QM/MM), 306, 307
- quantum mechanics (QM), 306
 potential, 317
- quasiharmonics, 318
- racemic coupling, 278*f*
- radical coupling, 73*f*
- Raman spectra, 200–203
 bleached spruce kraft pulp, 207*f*
Cladophora glomerata, 202*f*
 tunicate and *Valonia ventricosa*, 201*f*
- Raman spectroscopy, 189
- random/combinatorial chemistry, 274–5
- random coupling model, 269–72
- R&D. *See* research and development
- rDNA. *See* ribosomal DNA
- reaction wood tissues, 223–5, 224*f*
- recalcitrant cellulose, 426
 substrates, 418–20
- recombinant cellulolytic strategy, 488–9
- regular repeating model, 269–72
- renewable biological resource, abundant, 480
- research and development (R&D)
 catalytic fuels (mixed-alcohol) synthesis, 27
 cleanup and condition, 26–7
 economic viability, 19, 26–8
 feedstock and process interface, 26
 gasification studies, 26
 hydrolyzate conditioning and pretreatment needs, 20
 integration/demonstration needs, 27–8
 interface needs, 19–20
 process, 19–20, 19*f*, 22–3
 research barriers, 19*f*

- restraining potentials, 319
reversibly glycosylated polypeptides (RGPs), 133
RG I. *See* rhamnogalacturonan I
RG-II. *See* rhamnogalacturonan II
RG-II:ApiT. *See* rhamnogalacturonan II:apiosyltransferase
RG-II:AT. *See*
 rhamnogalacturonan-II:acetyltransferase
RG-II:GalAT. *See* rhamnogalacturonan-II:galacturonosyltransferase
RG-II:GlcAT. *See* rhamnogalacturonan-II:glucuronosyltransferase
RG-II:MT. *See*
 rhamnogalacturonan-II:methyltransferase
RG-II:XyIT. *See*
 rhamnogalacturonan-II:xylosyltransferase
rhamnogalacturonan I (RG-I)
 acetyltransferase (RG-I:AT), 130, 136
 arabinosyltransferase (RG-I:AraT), 133–5
 galactosyltransferase (RG-I:GalT), 132–3
 methyltransferase (RG-I:MT), 135–6
 synthesized, 130–32
rhamnogalacturonan II (RG-II)
 apiosyltransferase (RG-II:ApiT), 128
 galacturonosyltransferase (RG-II:GalAT), 128
 glucuronosyltransferase (RG-II:GlcAT), 129–30
 methyltransferase, (RG-II:MT), 130
 synthesis (RG-II), 128–32
 transferases, 130
 xylosyltransferase (RG-II:XyIT), 129
RHD1. *See* ROOT HAIR DEFICIENT
ribosomal DNA (rDNA), 467
ribosomal sequence tags (RSTs), 469–70
RISA. *See* rRNA, intergenic spacer analysis
RMSD. *See* root mean square distance
ROOT HAIR DEFICIENT (RHD1), 144
root mean square distance (RMSD), 319
root pathogens, 460
rosettes (Cels), 42
rRNA
 automated (ARISA), 467
 intergenic spacer analysis (RISA), 467
 sequencing and quantification, 467–8
RSTs. *See* ribosomal sequence tags
rumen, 393
 cellulolytic and hemicellulolytic bacteria from, 394–7
 protozoan, phylogenetic relationship, 399*f*
Rumen
 fungi, 398
 protozoa, 398–9
Ruminococcus, 400, 410
 albus, 396
 Ruminococcus flavefaciens, 394–6
 cellulosome structure, 395*f*
 saccharophagus degradans, 383
 S-adenosyl-L-methionine (SAM), 137
 Sakakibara model, 272–3
 SALK T-DNA, 159
 salvage pathway, 148–9
 SAM. *See* S-adenosyl-L-methionine
 saprophytic fungi, bacteria and, 463–4
 SARST. *See* serial analysis of ribosomal sequence tags
 Scaffoldin, 408
 selective thermochemical processing, 33*f*
 Selfish Glycan Theory, 157
 separate hydrolysis and fermentation (SHF), 481–2
 serial analysis of ribosomal sequence tags (SARST), 469
 SHF. *See* separate hydrolysis and fermentation
 SHG. *See* signals of harmonic generation
 signals of harmonic generation (SHG), second and third, 50
 Silurian period, 213
 SIMREL, 272
 simultaneous saccharification and co-fermentation (SSCF), 482
 sinapyl alcohol dehydrogenase, 233–4
 SLOPPY enzyme, 142–3
 sodium hydroxide pretreatment, 444
 soft rot, 459
 soil bacteria, 460
 solid state CNMR, 189
 solvents
 cellulose-dissolving, 446
 models, 314–15
 sorghum, 254
 spectra differences, 203
 spruce
 delignification, 271
 tracheid wall layer, 83*f*
 SSCF. *See* simultaneous saccharification and co-fermentation
 static electronic structure theory, 334–5
 steam explosion, 440–41
 steered MD, 319
 strain development, 31–2
 stretch energy, 308, 309*f*
 suberin domains, 78
 sucrose synthase (SUSY), 102, 143–4
 sugar
 activated biosynthesis, 140–42
 flux, 155–6
 kinase, 140
 supercritical fluid pretreatment, 446

- SUSY. *See* sucrose synthase
- synapyl alcohol dehydrogenase, 233–4
- synthetic biology, 4–5
- synthetic dehydropolymerizates (DHPs), 241
- system size limitations, 316–17
- systems biology, 30–32
 research process, 31*f*
- targeted MD (TMD), 319
- TBAF. *See* tetrabutylammonium fluoride
- T-DNA insert mutants, 123
- technology integration, 34–5
- terminal restriction fragment length
 polymorphism (T-REL), 467, 468
- tetrabutylammonium fluoride (TBAF), 236
- thermal transformation, selective, 32–4
- Thermobifida fusca*, 462
- thermochemical conversion technologies, 10*f*,
 11, 12
 options, 23–5
- thermochemical ethanol production process,
 25*f*
- thermodynamic cycle, 321*f*
- thermodynamic integration, 320–22
- thioacidolysis, 240–41
- thioglycolic acid, 239–40
- TIRF. *See* total internal reflection fluorescence
- TM. *See* transmembrane domains
- TMD. *See* targeted MD
- tobacco, 252–3
 CAD downregulation, 258
 CCR and, 255
 vascular integrity, 258–9
- torsional energy, 309–10
 variation, 310*f*
- total internal reflection fluorescence (TIRF),
 50–51
- tracheid, 82*f*
- transferases, 228–9
- transformation systems, 487–8
- transmembrane (TM) domains, 97
- transportation fuel costs, 481*f*
- T-REL. *See* terminal restriction fragment
 length polymorphism
- trichid cell surface structure, 46*f*
- Trichoderma reesei*, 3, 458*f*
 cellulase, 380–81
- tyrosine ammonia lyase, 227–8
- UAM. *See* UDP-arabinopyranose mutase
- UDP- β -D-glucose (UDP-Glc), 143–4
- UDP- β -D-glucuronic acid (UDP-GlcA),
 147–50
 formation, 149–50
- UDP-Api, 151
- UDP-Ara, 152–3
- UDP-arabinopyranose mutase (UAM), 133
- UDP-arabinose furanose, 153
- UDP-Gal, 146
- UDP-galacturonic acid, 150–51
- UDP-Glc. *See* UDP- β -D-glucose
- UDP-GlcA. *See* UDP- β -D-glucuronic acid
- UDP-GlcNAc-binding sites, 99
- UDP-Glc pyrophosphorylase (UGLCP), 143
- UDP-Rha, 146–7
- UDP-sugar pyrophosphorylase, 142–3
- UDP-Xyl, 151
- UGLCP. *See* UDP-Glc pyrophosphorylase
- Valonia*, 190
 cellulose degradation, 422*f*
 cellulose ribbon degradation, 419*f*
 recalcitrant cellulose, 42
- vascular integrity, 268
- vascular plant diversification, 218–23
- Waals forces, 311
- wall fiber formation, 268
- wall polymers
 covalent interactions, 71–8
 molecular associations, 70–8
 non-covalent interactions, 70–71
- water density, local, 319
- water molecules, solvent, 339–40
- weighted histogram analysis method (WHAM),
 322
- WHAM. *See* weighted histogram analysis
 method
- world oil price scenarios, 22*f*
- XG. *See* xyloglucan
- XGA. *See* xylogalacturonan
- XT1. *See* xyloglucan xylosyltransferases
- xylan, 108–10, 365–6
 removal, 449*f*
- xylan hydrolysis, 331–3, 345, 347
 enzyme participation, 358*f*
- xylanases, 359–60
- xylanolytic enzymes
 hemicellulose fraction, 359*t*
 synergy, 359*t*
- xylariaceous Ascomycota, 456
- xylobiose
 calculations, 340–43
 decomposition, 345*f*, 346*f*
 dehydration, 342*f*
 formation, 347*f*
 hydrolysis, 333–5, 345
- xylogalacturonan (XGA), 69
- xyloglucan, 64–6, 366–7

-
- xyloglucan (XG), 105–8
 - structure, letter code elements, 106*f*
 - synthesis, defining, 107
 - xyloglucan xylosyltransferases (XT1), 107
 - xyloglucanase, 361–2
 - activities, 460
 - xylo-oligomer
 - degradation, 331–3
 - direct dehydration, 348
 - xylose
 - dehydration, 342*f*
 - formation, 345*f*, 347*f*
 - molecule, rapid proton transfer, 339*f*
 - xylose degradation, 331–3
 - CPMD simulation, 337*f*
 - quantum mechanical calculations, 333–5
 - vacuum reactions, 335–9
 - xylose reaction
 - proton affinities and activation energies, 338*f*
 - protonated reactions, 336*f*
 - Yellowstone National Park hot spring, 461–2